Guckelsberger, Christian


Published: 30/11/2020

Document Version
Publisher's PDF, also known as Version of record

Please cite the original version:
INTRINSIC MOTIVATION IN COMPUTATIONAL CREATIVITY APPLIED TO VIDEOGAMES

CHRISTIAN GUCKELSBERGER
OCTOBER 2020

A thesis submitted in partial fulfilment of the requirements for the degree of
Doctor of Philosophy

Examiners:
Geraint A. Wiggins (Queen Mary, Vrije Universiteit Brussel),
Georgios N. Yannakakis (University of Malta)

Supervisors:
Simon Colton (Queen Mary, Monash University),
Jeremy Gow (Queen Mary), Paul Cairns (University of York),
Christoph Salge (University of Hertfordshire, New York University)

This work was supported by the EPSRC doctoral training programme
Intelligent Games - Game Intelligence (IGGI, grant no. EP/L015846/1).

It was carried out in the Computational Creativity Group, Department of
Computing, Goldsmiths, University of London, and in the Game AI Research
Group, School of Electronic Engineering and Computer Science, Queen Mary,
University of London. Further work was conducted during a research visit at the
Game Innovation Lab at New York University, Brooklyn, New York.
Truisms have the disadvantage that by dulling the senses they obscure the truth. Almost nobody will become alarmed when told that in times of continuity the future equals the past. Only a few will become aware that from this follows that in times of socio-cultural change the future will not be like the past. Moreover, with a future not clearly perceived, we do not know how to act with only one certainty left: if we don’t act ourselves, we shall be acted upon. Thus, if we wish to be subjects, rather than objects, what we see now, that is, our perception, must be foresight rather than hindsight.


The Ethical Imperative: Act always so as to increase the number of choices. The Aesthetical Imperative: If you desire to see, learn how to act.

STATEMENT OF ORIGINALITY

I, Christian Guckelsberger, confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged in the following and my contribution indicated. Previously published material is also acknowledged in the following section.

I attest that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge break any UK law, infringe any third party’s copyright or other Intellectual Property Right, or contain any confidential material.

I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis.

I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university.

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without the prior written consent of the author.

Helsinki, 23rd October 2020

__________________________________________
Christian Guckelsberger
Over the time of this PhD, 13 conference and journal papers have been published that relate directly to this final thesis, with ideas, results and figures featuring in subsequent chapters. These publications are listed below, with details on the contributions of the respective co-authors.

2016


*CG and CS jointly devised the formalism. RS informed on related work. CG developed the thought-experiments with input from CS and SC. CG wrote the article. CS and SC provided comments on earlier drafts.*


*CS proposed the study. CG compiled the theoretical argument and wrote the article. A prior version was submitted to the 'Introduction to the Philosophy of Cognitive Science’ class at University College London in Fall 2015. The class was hosted by David Lagruado, who provided additional feedback. Furthermore, Martin Biehl and Janet Gibbs provided helpful comments on draft versions.*


*CG and CS motivated and designed the study. CG surveyed related work, conducted the experiments and introduced modifications to the formalism. SC provided feedback on early drafts. CG wrote the article, and Jeremy Gow provided feedback.*

2017


*CG and CS motivated the investigation. CG and PC designed the study with further feedback from CS and JG. CG conducted and evaluated the study and wrote the article. JG provided further feedback on earlier drafts.*

CG and SC conceived of the motivation and thought-experiment. CG and CS finished the theoretical argument. CG wrote the article.


AD and DZ conceived of the survey. CG added related work, provided feedback on the conceptualisation of challenge and contributed to the write-up.

2018


SR, JT and PH produced an initial draft of the survey. CG added a substantial amount of further related work and contributed to the write-up.


CS and TM designed the approximation and conducted a preliminary study. CS, CG and RC extended the study. CS conducted and evaluated the experiments. CS, CG and RC wrote the article.


CG and CS motivated the study and extended the formalism. CG, CS and JT contributed to the study design. CG conducted the study and wrote the article.


MB, CG, CS, SS, and DP conceived of the study, discussed the concepts, revised the formal analysis, and wrote the article. MB contributed the initial formal analysis.

2020

AD and DZ conducted the survey. AD and PC did the statistical analysis. CG contributed to the conceptualisation of challenge, related work and the write-up.


SL, CG and AK conceived of the study, devised the formalism, and wrote the paper.


SC and AP conceived of the idea. CS, AP, CG, JM and MTL contributed to its elaboration. All contributed to writing the paper.
Computational creativity (CC) seeks to endow artificial systems with creativity. Although human creativity is known to be substantially driven by intrinsic motivation (IM), most CC systems are extrinsically motivated. This restricts their actual and perceived creativity and autonomy, and consequently their benefit to people. In this thesis, we demonstrate, via theoretical arguments and through applications in videogame AI, that computational intrinsic reward and models of IM can advance core CC goals.

We introduce a definition of IM to contextualise related work. Via two systematic reviews, we develop typologies of the benefits and applications of intrinsic reward and IM models in CC and game AI. Our reviews highlight that related work is limited to few reward types and motivations, and we thus investigate the usage of empowerment, a little studied, information-theoretic intrinsic reward, in two novel models applied to game AI.

We define coupled empowerment maximisation (CEM), a social IM model, to enable general co-creative agents that support or challenge their partner through emergent behaviours. Via two qualitative, observational vignette studies on a custom-made videogame, we explore CEM’s ability to drive general and believable companion and adversary non-player characters which respond creatively to changes in their abilities and the game world.

We moreover propose to leverage intrinsic reward to estimate people’s experience of interactive artefacts in an autonomous fashion. We instantiate this proposal in empowerment-based player experience prediction (EBPXP) and apply it to videogame procedural content generation. By analysing think-aloud data from an experiential vignette study on a dedicated game, we identify several experiences that EBPXP could predict.

Our typologies serve as inspiration and reference for CC and game AI researchers to harness the benefits of IM in their work. Our new models can increase the generality, autonomy and creativity of next-generation videogame AI, and of CC systems in other domains.
ACKNOWLEDGMENTS

This thesis rests on the support of many, and what follows is an incomplete list of the people who helped me so profoundly along the way.

I am lucky to have found an extraordinary team of supervisors who complemented each other greatly. Many thanks to Simon for thought-provoking (and heated) discussions on computational creativity, his invaluable advice on succeeding in academia, and for teaching me British manners. I am indebted to Jeremy for his brilliant mind and selfless support. I thank Paul for his comforting tips on staying healthy and sane during the PhD marathon, and for his bullet-proof help with the design and evaluation of experiments. Finally, many thanks to Christoph for countless hours of in-depth supervision and dazzling discussions on intrinsic motivation, information theory and philosophy. With Georgios and Geraint, I have won the best possible examiners for this thesis. Thank you so much for providing me with excellent feedback, and for turning my viva into a truly rewarding experience.

This work has immensely benefited from conversations with colleagues and friends. Daniel Polani has been a devoted academic mentor, and an indispensable source of knowledge on information theory and empowerment. I am grateful to Martin Biehl for his feedback and discussions on the intricacies of intrinsic motivation. I want to extend my gratitude to Anna Jordanous, Anna Kantosalo and Simo Linkola for letting me pick their brain on evaluating computational creativity. Moreover, I thank Mike Cook for insights on automatic game design, Richard Bartle for help with game design terminology, and Rob Saunders for his thoughts on intrinsic motivation in computational creativity. Many thanks to Julian Togelius and Hannu Toivonen for supporting extensive research visits to the Game Innovation Lab at New York University and the Discovery Research Group at the University of Helsinki.

I am indebted to my family and in particular my parents for their eternal faith in my abilities, and for supporting me throughout this project. I am incredibly grateful for Reingard and Michael’s heart-warming encouragement and advice based on their own PhD experience. Also, I cannot thank Peter and Constanza enough for their friendship, and for giving me a home away from home. I am grateful to my friends in Hamburg for turning countless days in the library into such warm, rewarding experiences. The last months towards submission were particularly tough; my thanks go to all those dear friends who encouraged me in these times, and especially to Sophia, Theo, Dizzy and Alex for giving me shelter. Finally, many thanks to my friends and colleagues within and beyond the IGGI programme for sharing this journey with me, in particular to Tom, Daniel and Memo for being such inspiring personalities, and for making the years at Goldsmiths truly memorable.

Finally, I am incredibly grateful to Marlene, who has supported me throughout the past five years like nobody else could, and, while putting up with many sacrifices along the way, always encouraged me to stay on course.

Helsinki, 23rd October 2020

Christian Guckelsberger
8.4 Consolidation ........................................... 250

9 CONCLUSION .............................................. 252
  9.1 Research Questions Revisited ........................... 253
  9.2 Contributions and Potential Impact ..................... 256
  9.3 Concluding Remarks .................................... 258

V APPENDICES: MATHEMATICAL FOUNDATIONS
  A PROBABILITY THEORY ................................... 261
  B CAUSAL BAYESIAN NETWORKS ......................... 265
  C INFORMATION THEORY .................................. 270
  D PERCEPTION-ACTION LOOP ............................... 276

BIBLIOGRAPHY ................................................. 281
| Figure 3.1 | Generative model of perception-action-loop | 49 |
| Figure 3.2 | Empowerment action-value in the generative model | 54 |
| Figure 3.3 | Empowerment for deterministic dynamics | 58 |
| Figure 3.4 | Empowerment for stochastic dynamics | 60 |
| Figure 3.5 | Empowerment maximisation in a gridworld | 62 |
| Figure 3.6 | Empowerment maximising pendulum | 63 |
| Figure 4.1 | Algorithmic vs. heuristic task continuum | 82 |
| Figure 4.2 | Typologies of the benefits and applications of intrinsic reward and motivation in computational creativity | 87 |
| Figure 5.1 | Multi-agent perception-action-loop | 154 |
| Figure 5.2 | Multi-agent perception-action loop generative model | 155 |
| Figure 5.3 | Coupled empowerment in generative model | 161 |
| Figure 5.4 | CEM-induced operational proximity | 174 |
| Figure 5.5 | Trust in coupled empowerment | 176 |
| Figure 5.6 | CEM-induced protection of the player | 178 |
| Figure 5.7 | Health-performance consistency for player protection | 179 |
| Figure 5.8 | Player attraction via transfer empowerment | 180 |
| Figure 5.9 | CEM-driven NPC rescuing player and themselves | 181 |
| Figure 5.10 | Empowerment variants in predator-and-prey scenario | 187 |
| Figure 5.11 | CEM-driven daredevil predator | 188 |
| Figure 5.12 | NPC empowerment for embodiment modifications | 189 |
| Figure 5.13 | CEM-induced lava pushing | 190 |
| Figure 5.14 | NPC-player transfer empowerment in remote attack | 191 |
| Figure 5.15 | CEM-induced remote attack | 192 |
| Figure 6.1 | Expressive range in No Man’s Sky | 204 |
| Figure 6.2 | Constructive PCG in The Division Underground | 206 |
| Figure 6.3 | RoboRunner game testbed | 218 |
| Figure 6.4 | Procedural content generator for RoboRunner levels | 219 |
| Figure 6.5 | Genotype-phenotype mapping of RoboRunner levels | 219 |
| Figure 6.6 | Individual steps in empowerment-based player experience prediction for a RoboRunner level | 221 |
| Figure 6.7 | Procedurally generated RoboRunner study conditions | 222 |
| Figure 6.8 | CEM-based antagonism in Curious Whispers | 244 |
| Figure 6.9 | CEM-based support in Curious Whispers | 244 |
| Figure 6.10 | Player experience prediction in the GVGAI framework | 246 |
| Figure B.1 | Bayesian network | 266 |
| Figure B.2 | I-equivalent Bayesian network | 267 |
| Figure B.3 | Intervention in a causal Bayesian network | 268 |
| Figure C.1 | Communication system | 271 |
| Figure C.2 | Entropy of a binary random variable | 272 |
| Figure C.3 | Venn diagram of information-theoretic quantities | 274 |
| Figure D.1 | Perception-action-loop | 277 |
LIST OF TABLES

Table 4.1 Reviewed work using intrinsic reward or intrinsic motivation in computational creativity . . . . . . . . . . . . 86
Table 5.1 Reviewed work using intrinsic reward or intrinsic motivation in videogame AI . . . . . . . . . . . . . . . . . . . . 118
Table 6.1 Determinants of believable NPC behaviour . . . . . 144
Table 6.2 Level elements in our study of companion NPCs . . 172
Table 6.3 Character abilities in our study of companion NPCs . 173
Table 6.4 Level elements in our study of adversary NPCs . . . 184
Table 6.5 Character abilities in our study of adversary NPCs . . 185

LIST OF ALGORITHMS

Algorithm 1 Coupled empowerment maximisation . . . . . 167
Algorithm 2 Empowerment-based player experience prediction . 214

LIST OF ACRONYMS

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AI</td>
<td>Artificial intelligence</td>
</tr>
<tr>
<td>BN</td>
<td>Bayesian network</td>
</tr>
<tr>
<td>CC</td>
<td>Computational creativity</td>
</tr>
<tr>
<td>CEM</td>
<td>Coupled empowerment maximisation</td>
</tr>
<tr>
<td>EBPXP</td>
<td>Empowerment-based player experience prediction</td>
</tr>
<tr>
<td>EM</td>
<td>Empowerment maximisation</td>
</tr>
<tr>
<td>GVGAI</td>
<td>General video game AI</td>
</tr>
<tr>
<td>HCI</td>
<td>Human-computer interaction</td>
</tr>
<tr>
<td>IM</td>
<td>Intrinsic motivation</td>
</tr>
<tr>
<td>IR</td>
<td>Intrinsic reward</td>
</tr>
<tr>
<td>IMRL</td>
<td>Intrinsically motivated reinforcement learning</td>
</tr>
<tr>
<td>NPC</td>
<td>Non-player character</td>
</tr>
<tr>
<td>PA</td>
<td>Perception-action</td>
</tr>
<tr>
<td>PCG</td>
<td>Procedural content generation</td>
</tr>
<tr>
<td>PX</td>
<td>Player experience</td>
</tr>
<tr>
<td>RL</td>
<td>Reinforcement learning</td>
</tr>
</tbody>
</table>
INTRODUCTION

Can we engineer artificial systems that are creative? This is one of the central questions of computational creativity (CC), a subfield of artificial intelligence (AI) and the backdrop of this thesis. The challenge to endow artificial systems with creativity has accompanied the development of AI since its inception (cf. McCarthy et al., 1955/2006), and is regarded by some as its final frontier (Colton & Wiggins, 2012): not only does creativity draw on many faculties of intelligence, but it might also be necessary for artificial general intelligence. To be perceived as intelligent across different situations, and in the absence of a solution or even a specific problem (cf. Stanley, Lehman & Soros, 2017; Wiggins, 2018), an artificial systems may have to exhibit behaviours and generate products that would be commonly considered creative.

CC researchers draw on insights from a wide range of disciplines such as cognitive science, engineering, mathematics, philosophy, sociology and art (McGregor, Wiggins & Purver, 2014; Ackerman et al., 2017) to further a twofold research agenda (Pérez y Pérez, 2018; Veale, Cardoso & Pérez y Pérez, 2019). On the one hand, they generate insights into the phenomenon of (human) creativity by means of computational modelling and empirical studies. On the other hand, they engineer systems which embody these insights to benefit people culturally, socially and economically (Loughran & O’Neill, 2017; Smith, 2017). This requires artificial systems to take over at least some creative responsibility in different tasks (Colton & Wiggins, 2012). A central challenge in this endeavour is to increase a system’s creative autonomy (Jennings, 2010; Saunders, 2012; McCormack, Gifford & Hutchings, 2019), i.e. to reduce their reliance on people during creative activity in order to be ultimately considered creative in their own right (Colton, 2008).

Both parts of this agenda have been pursued for different types of creativity (Kaufman & Beghetto, 2009), from everyday creative acts to eminent creative achievements. Moreover, CC research has been conducted in a large range of creative domains (Loughran & O’Neill, 2017) such as problem solving, mathematical theory formation, scientific discovery, graphic design, audio design, the visual arts, the culinary arts, sculpture, choreography, musical composition, musical accompaniment, fictional ideation and various creative language applications, such as poetry, storytelling, narrative design, neologisms, metaphor, slogans, etc. Traditionally, CC has focussed on the production of artistic artefacts (Colton & Wiggins, 2012), rather than on creative behaviour or creativity in problem-solving (Besold, 2016).

Amongst all creative domains, videogames could be considered the drosophila of CC. Games invite research on many types of creativity in both play (Zook, Riedl & Magerko, 2011; Liapis, Yannakakis & Togelius, 2014; Moffat, 2015) and design (Liapis, Yannakakis & Togelius, 2014; Ventura, 2016a). As a popular benchmark for artificial general intelligence, games facilitate the comparison and exchange of findings with CC. They bring together different content facets in a complex whole, some of which are already being investigated in
1.1 Motivation

CC has been and continues to be substantially shaped by the work of Boden (1992, 1990/2003), who put forward a theoretical framework for the study of creativity in AI. Boden describes the creative process in terms of different creative mechanisms, but she crucially does not address its very foundation: a person’s or artificial agent’s motivation as a driving force of their agency (Saunders, 2012). Today, CC researchers have filled this theoretical gap with different specific action-selection mechanisms; however, they rarely refer to an agent’s motivation as a means to distinguish and compare these formalisms.

This is not surprising, as the majority of existing CC systems only realise one type of motivation which psychologists call extrinsic: they choose actions to satisfy specific, external goals¹ of their users or designers. Even if a system appears to follow their own internal goals, their creative activity is a means to an end. This also characterises many instances of human creativity, e.g. in the workplace: we perform creatively on tasks which eventually yield a separate, instrumental value for the company and ourselves, e.g. in the form of revenue and a wage. However, this work is typically not rewarding in itself.

¹ These goals may be externally communicated, or internalised through a pre-defined evaluation function which can be hard-coded or trained on externally provided data.
While extrinsic motivation dominates the CC research landscape, it crucially does not fully explain the many facets of human creativity. From small, everyday creative acts to major scientific or artistic achievements, the human creative process is often guided by inherent satisfactions, rather than a separate consequence (Ryan & Deci, 2000a). Technically speaking, much of human creativity is intrinsically motivated. One of the most prominent examples of such intrinsic motivation (IM) is curiosity: our creative work as researchers, for instance, is frequently driven by the desire to learn something new. Similarly, people are more likely to push the boundaries of a mathematical theory, or to discover a new style of painting, through acts of curiosity that are rewarding in themselves, rather than for separate, potentially externally dictated, goals. The central role of IM as a core mechanism of individual human creativity is supported by decades of empirical work in psychology (Amabile, 1983; Amabile & Pillemer, 2012; de Jesus et al., 2013; Liu et al., 2016).

We believe that CC’s reliance on extrinsic motivation hinders progress on both ends of its research agenda. CC researchers can generate only limited insights into the nature of human creativity through computational modelling if they do not account for the important role of IM. Moreover, the present motivational focus might counteract the goal to build systems that unbiased observers deem creative (Colton & Wiggins, 2012). The philosopher Kieran (2014) observes that ‘we tend to be wary of those [people] whose goals concerning [a creative] activity are stated in purely extrinsic terms’ (ibid., p. 14). Similarly, we expect people to perceive extrinsically motivated CC systems, assuming that these communicate the reasons for their actions (Charnley, Pease & Colton, 2012), as less creative. Extrinsic motivation may not only influence the perception of creativity (Colton, 2008), but also limit creativity itself: studies in psychology show that many types of extrinsic reward have a detrimental effect on people’s creativity (Amabile, 1998; Malik & Butt, 2017; Amabile, 2018). The same may hold for artificial systems, especially if their design is inspired by human cognition. Putting psychological parallels aside, formal limitations remain: an extrinsically motivated system relies on its user or designer to fixate often context-specific, separate goals. This intuitively counteracts the system’s creative autonomy and the central CC goal to engineer systems that can be considered creative in their own right (Colton, 2008).

Crucially though, AI researchers have adopted the concept of IM from human psychology, and formalised it in computational models (Oudeyer & Kaplan, 2007). These models drive action-selection through the optimisation of a formal intrinsic reward (IR). Some researchers have done pioneering work by investigating specific IM models in CC, both theoretically and in applied studies. However, they have not related the effects of these specific models to IM as a larger family of motivational mechanisms. In fact, most related work does not even refer to the notion of ‘intrinsic motivation’. Until now, it has consequently remained unclear how models of IM more generally could benefit CC. Yet, the preliminary results on specific models are promising: they suggest that IM could overcome the drawbacks of extrinsic motivation, most prominently its limitations to a system’s creativity and creative autonomy.

As CC ‘killer applications’ (Liapis, Yannakakis & Togelius, 2014, p. 1), videogames represent a particularly suitable domain to yield insights into
CC more generally through applications of IM models. Moreover, games are quintessential intrinsically motivating activities (Salen & Zimmerman, 2004, p. 333), and we thus expect game AI to be particularly receptive to the use and advantages of IR and IM models. Existing work suggests that IM can substantially improve the robustness and generality of AI applied to videogame tasks, e.g. in driving the behaviour of game-playing agents or in procedurally generating game content. However, this work has again mostly considered specific models, and, if considered in isolation, tells us little about the benefits of IM models for game AI more generally.

The aim of this thesis is to understand and demonstrate how IM in general, as a distinct class and as a family of motivational mechanisms, can advance CC and game AI, and which challenges researchers must face. We support our interdisciplinary, big picture view by new models and applications of IM to game AI. To this end, we adopt and extend empowerment and empowerment maximisation (EM) as specific intrinsic reward and intrinsic motivation, respectively. Informally, empowerment quantifies an agent’s potential and perceivable influence on their environment, including other agents, and EM yields agents that maximise their options and control. We argue that IM more generally and EM specifically have the potential to transform the game engineering and design practice, as well as the experiences of players.

1.2 RESEARCH QUESTIONS

The overarching research questions (RQs) of this thesis are:

RQ.1 Can IR and models of IM advance CC?

RQ.2 Can IR and models of IM advance videogame AI?

These questions have neither been answered yet, nor are they easy to answer. They are not readily supported by existing work, as it has been conducted in isolation and does not relate to an overarching, unifying concept of IM. Moreover, a strong affirmation requires more than a single example; it necessitates a deep exploration of the meaning of IR and IM, their beneficial properties, and potential applications across both domains. We demarcate the fields of CC and game AI at this level to allow members of either research community to identify and retrieve what is of immediate relevance to them. We qualify these overarching questions separately for each discipline with specific research questions in our contribution chapters:

RQ.3 Why have IR and models of IM been used in CC?

RQ.4 How have IR and models of IM been used in CC?

We have raised these questions to support RQ.1 through existing research. They ask for which reasons, in the sense of properties, and for which CC applications, i.e. to which end, IR and IM models have been used in the past. We direct similar questions to existing work in game AI to support RQ.2:

RQ.5 Why have IR and models of IM been used in videogame AI?

RQ.6 How have IR and models of IM been used in videogame AI?
We serve the computational game creativity research agenda by asking whether applications of IR and IM in game AI and CC can be related to each other. The insights allow for answers to RQ.1 to support RQ.2 and vice versa:

RQ.7 How do existing applications of IR and IM models in videogame AI and CC overlap?

We complement these retrospective questions by asking whether IR and models of IM can benefit CC through two novel applications. We demonstrate the diverse uses of IR and IM models by addressing generation and evaluation as the core components of the creative process, across the two domains of simple creative behaviours and complex artefacts:

RQ.8 Can we use a model of intrinsic motivation to engineer general and social co-creative agents?

RQ.9 Can we use IR to predict people’s experience of interactive artefacts in a general and autonomous way?

We answer these questions by proposing an informal solution for the CC scenario, and then formalising it to tackle a game AI challenge which instantiates this scenario. We address RQ.8 by investigating whether a model of IM can be used to drive the behaviour of general, believable non-player characters that either support a player as companions, or challenge them as adversaries. To realise different types of creativity, we particularly rely on the capacity of IM to yield emergent behaviour. Moreover, we address RQ.9 by examining whether IR can be used to predict players’ experiences of procedurally generated content independently of players or designers.

We thus study CC ‘within and for computer games’ (Liapis, Yannakakis & Togelius, 2014, p. 2), supporting RQ.1 through RQ.2. Our new models are informed by existing work in both domains, and we hence also realise Ventura’s (2016) proposition to shape the future of game AI through CC research, supporting RQ.2 through answers to RQ.1.

Our research aim takes us into largely uncharted territory, and we have hence formulated these questions in an open-ended way. To explore them efficiently, we exclusively rely on studies that yield rich qualitative data.

1.3 CONTRIBUTIONS

We finish this introduction by summarising our main contributions to CC, game AI, and AI more generally. We relate them to our research questions, and highlight the structure of this thesis along the way. In the tradition of CC research, we draw on many other fields, in particular psychology, philosophy and AI. To make this thesis accessible to researchers across all these fields, we introduce key theoretical and mathematical concepts from scratch. In particular, Appxs. A – D serve as reference to the mathematical foundations. We lay out core concepts in probability and information theory for a common treatment of artificial and biological agents as information processing systems. This culminates in the formalism of the perception-action (PA)-loop, which we use throughout this thesis to describe the interaction of an agent with their surroundings over time, and to quantify the information flows within.
In Ch. 2, we introduce the reader to the concepts of IR and IM, and to their formalisation in AI. We describe how IM emerged as a concept in psychology, and discuss the most relevant theories as explanations of intrinsically motivated behaviour in people and as inspiration of many computational models. We highlight the incentives of AI researchers to formalise IR and IM, and summarise interdisciplinary efforts in defining IM more accurately. Informed by this debate, we put forward a working definition of IM models. We introduce well-acknowledged models of IM, negative examples, and debatable cases to test our definition and to demonstrate the breadth of IM models to the reader. Our working definition is a prerequisite to distinguish and relate intrinsic and extrinsic motivational models in existing CC and game AI work in later chapters. The main contributions of this chapter are:

- An extensive, interdisciplinary account of IR IM in psychology and AI as means to foster mutual understanding between both disciplines.
- A tested, informal working definition of IM models, based on four diagnostics of an IR function embedded in a motivational model.

In Ch. 3, we introduce and motivate empowerment as well as EM (Klyubin, Polani & Nehaniv, 2005b; Salge, Glackin & Polani, 2014b), the formal IR and IM to be adopted and extended in this thesis. We provide a generic and a simplified formalisation of empowerment and EM that distinguishes between an agent’s objective world and their beliefs about that world, and thus overcomes ambiguity in prior work. To support the reader’s intuitions, we illustrate properties of empowerment and provide examples of the behaviour of an empowerment maximising agent. Moreover, we sketch the parts of the empowerment research landscapes that we contribute to through our applied work in Ch. 5 and 7. The main contributions of this chapter are:

- An updated introduction to empowerment and EM in discrete scenarios. It comprises the strongest motivation of the intrinsic reward and motivational principle to date, drawing on more insights from psychology, biology and physics than any instance of related work.
- A generic formalisation of empowerment and EM, which distinguishes an agent’s objective world and their beliefs about that world, and thus makes implicit assumptions in prior work transparent. We moreover simplify this formalisation considerably for application in our models.

These background chapters lay the foundations for the second part of this thesis, where we map the landscape of existing work on IR and IM in both domains, and for the first time bring it together under one umbrella by leveraging our working definition of IM.

In Ch. 4, we present a comprehensive, systematic review to identify why (RQ.3) and how (RQ.4) IR and models of IM have been leveraged in existing CC work. We draw on creativity studies to highlight the challenge of defining creativity, and to clarify our position in this thesis. We introduce the reader to CC, and, based on a critical survey of definitions and dominant positions, formulate a more inclusive working definition of CC that warrants the inclusion of related work from a wider range of AI subfields. By applying this
definition and our working definition of IM (Ch. 2), we retrieve 28 related work items from 1998 to 2018. Based on an in-depth analysis of this work, we identify four properties of IR, two corollaries and four properties of intrinsically motivated behaviour as reasons to embrace IM in CC. Informed by (computational) creativity theories and empirical findings from creativity studies, we moreover extract 12 (abstract) applications of IM to CC. We map and connect these properties and applications in two typologies. Based on our new insights, we motivate and contextualise our applied contributions. The main contributions of this chapter are:

- The first systematic review of IM in CC. Related work is classified in terms of the creative domain, system details (e.g. autonomous/co-creative, single/multi-agent, etc.) and the IM model used.

- Two typologies of the reasons to embrace IM in CC, and the (abstract) applications of IR and IM models to CC.

In Ch. 5, we complement the previous study with a systematic review of existing work using IM in game AI, as answer to RQ.5, RQ.6 and RQ.7. We inform this by game design and games user research findings on what makes games intrinsically motivating for people. We retrieve 11 game AI studies from 2006 to 2019 via our working definition of IM models (Ch. 2), Yannakakis and Togelius’s (2018, p. 4) definition of game AI, and Juul’s (2003) definition of videogames. We bias our selection towards studies that aim to benefit games, rather than using games as a benchmark for artificial general intelligence. We crucially uncover the same reasons to embrace IR and models of IM identified in Ch. 4. Leveraging the insights from game design and games user research, we develop a typology of 11 (abstract) applications of IM across four core domains of videogame AI. We answer RQ.7 by identifying a strong overlap between applications of IR and IM in CC and game AI, which allows us to understand related game AI work as instances of computational game creativity. We again use these insights to motivate and contextualise our applied contributions. The main contributions of this chapter are:

- The first dedicated, systematic review of IM in game AI with a focus on how IR and IM models can benefit game engineers, designers and players. Related work is classified in terms of the game AI domains studied and the IM models used, amongst others.

- A typology of (abstract) applications of IR and IM models in the four core areas of videogame AI.

Our reviews support the identification of unexplored areas of inquiry. Our typologies can be used by researchers from both fields as inspiration and reference to harness the benefits of IM in their research. In the third part of this thesis, we take the identified, existing work further in two new applications of IR and IM models to game AI with benefits to CC more generally.

In Ch. 6, we raise and examine our question RQ.8 by identifying a need for more general, co-creative artificial agents that can support or challenge their partner in human-computer co-creativity. We argue that specific interactions of a player and a non-player character (NPC) constitute co-creative acts, and thus
frame NPC AI as a CC application. We consequently identify the analogue game AI challenge to drive the behaviour of general and believable NPCs that either support or challenge the player as companions and adversaries, respectively. We define social models of intrinsic motivation as a blueprint for IM models capable of realising such behaviour in NPC AI and human-computer co-creativity more generally. We argue that empowerment as IR can give rise to general, believable supportive and adversarial behaviour in videogames, and instantiate coupled empowerment maximisation (CEM) as a specific social IM model. We present two qualitative studies based on observational vignettes to probe this ability on a custom-made videogame. Our operationalisations of believable NPC behaviour are informed by game studies and games user research. We find, amongst others, that CEM as the underlying principle can yield both emergent supportive and adversarial behaviour that surprised even the experimenters. The NPCs exhibit generality by responding with new but sensible behaviours to changes in their environment and abilities, and we hence deem them creative. We thus support RQ.8 through an application in game AI. The main contributions of this chapter are:

- Social models of IM as a generic approach to yield supportive or adversarial agent behaviour in open-ended interaction.

- Coupled empowerment maximisation (CEM) as a specific social IM model. The generic and simplified formalisation entails the definition of transfer empowerment as a novel, social IR.

- Two extensive qualitative studies using observational vignettes to probe the capacity of CEM to drive general NPCs that either support or challenge the player as companions or adversaries.

In Ch. 7, we formulate and evaluate RQ.9 by identifying a shortcoming in existing CC approaches to evaluating people's experience of interactive artefacts: whenever the artefact or generator changes, people must be involved again to update the evaluation function. This inflexibility endangers the (perception of) creativity and the autonomy of CC systems. We retrieve the same challenge in the modelling of player experience (PX) in videogames, with a particularly detrimental impact on the expressive potential of procedural content generation (PCG). We address this challenge by proposing a new approach to predicting PX with IR calculated on simulated gameplay trajectories. Drawing on findings in game design, games user research and human-computer interaction, we identify empowerment as a potential PX predictor, and instantiate our generic proposal in empowerment-based player experience prediction (EBPXP). We explore which experiences EBPXP could predict through a qualitative experiential vignette study. More specifically, we conduct a thematic analysis on think-alouds of participants playing procedurally generated levels of a dedicated game. We find that levels which are predicted to yield different PXs by EBPXP indeed evoke different experiences in players, with the strongest effect on perceived challenge. This supports our RQ.9. By consulting related games user research, we shape the hypothesis that empowerment allows for the prediction of the foundational experiences of effectance, outcome uncertainty, and perceived control, and only has mediating effects on experiences such as challenge. The main contributions of this chapter are:
1.3 Contributions

- **Intrinsic reward-based player experience prediction** as a generic approach to modelling player experience independently of player feedback and designer knowledge about a game’s semantics.

- Empowerment-based player experience prediction (EBPXP) as an instantiation of this approach based on empowerment as IR.

- An exploratory study based on a procedurally generated experiential vignette to identify player experiences which EBPXP could predict.

This concludes our applied contributions and the main part of the thesis.

In Ch. 8, we highlight promising future directions to advance our four clusters of contributions, and report work in progress. We consider next steps on advancing our individual studies and, where applicable, discuss extensions to the underlying principles and their potential application to advance CC beyond game AI. We finally promote how our individual contributions can be consolidated to generate new insights within and beyond game AI.

In Ch. 9, we revisit our specific research questions and discuss our answers with respect to our overarching research aim. We consider how our core contributions could impact academic and industrial research beyond this work, and wrap up this thesis with concluding remarks.
Part I

BACKGROUND
The concept of intrinsic motivation (IM) is at the centre of this thesis. But what is IM, precisely? Why would we want to formalise and employ it in artificial agents? And how can we distinguish a model of IM from other motivational mechanisms in AI? In this chapter, we answer each of these questions in detail, and with regards to the state of the art in psychology and AI. These preparations allow us to better grasp the meaning of IM computational creativity (CC) and videogame AI, as discussed in Ch. 4 and Ch. 5. Moreover, they provide the foundation for a critical assessment of the fit and the leverage of our contributions in these domains.

In Sec. 2.1 we look at the origins of IM in psychology. We show how IM emerged as a new concept from the traditional view of motivation, introduce the reader to key distinctions, and discuss the most common theories put forward to explain intrinsically motivated behaviour. This foundational work in psychology has been of great inspiration to AI researchers, whose efforts to formalise IM into computational models are the subject of Sec. 2.2. We contextualise the incentives to formalise IM, and discuss the cross-disciplinary quest to define the concept. This informs our working definition of IM models, an essential tool to distinguish them from their extrinsic counterparts in our systematic reviews in Ch. 4 and 5, and to consider work on specific IM models under one umbrella. We establish further reference points and probe the quality of our definition by evaluating it against well acknowledged models of intrinsic and extrinsic motivation. These examples are kept informal to allow for a more straight-forward comparison to existing work.

The main contributions of this chapter are twofold. We present a comprehensive, interdisciplinary account of how IM is understood across both psychology and AI. Based on pioneering work and the latest state of research, we show how psychological theories have inspired formalisations of IM in AI, and vice versa how AI researchers’ need for a formal definition of IM, combined with subject-specific methodology, contributes to a better understanding of the concept across the disciplines. We synthesise these insights in our second main contribution: a working definition of IM models which could benefit researchers in a wide range of disciplines. This comprehensive approach is an essential ingredient for this interdisciplinary thesis, and we hold on to it throughout the next chapters.

2.1 ORIGINS AND THEORIES IN PSYCHOLOGY

The IM concept has its roots in psychology, where it is now widely studied and applied in the fields of educational, developmental and organisational psychology. In Sec. 2.1.1, we describe how the new concept originated as a response to observations that were irreconcilable with existing theories of motivation. We discuss how it has been defined, and why ‘intrinsic’ and ‘extrinsic’ are different qualifiers than ‘internal’ and ‘external’. In Sec. 2.1.2,
we provide insights on how IM is commonly operationalised, and outline the evolution of the most common psychological theories. This equips us with the basis for understanding why AI researchers became interested in the concept, and to recognise when and how models of IM draw on their psychological foundations.

2.1.1 Intrinsic vs. Extrinsic Motivation

To be motivated means ‘to be moved to do something’ (Ryan & Deci, 2000a, p. 54), i.e. motivation drives behaviour in terms of the selection of actions. Originally, psychologists only differentiated levels of motivation, characterising how much an organism is moved to act. Based on the behaviourist theory of operant conditioning (Skinner, 1953), action-taking had been related to a separable consequence or instrumental outcome. For instance, you might be reading this thesis as a means to collect material for your own writing.

This status quo was challenged by experimental studies of animals (Harlow, 1950; White, 1959), exhibiting ‘exploratory, playful, and curiosity-driven behaviors’ (Ryan & Deci, 2000a, p. 56) that could not be explained by such a separable consequence. Psychologists thus acknowledged that there must be two distinct types of motivation: An organism is considered intrinsically motivated (Harlow, 1950), if they engage in an activity because it is inherently interesting, enjoyable and satisfying. IM is volitional, and usually ‘accompanied by the experience of freedom and autonomy’ (Ryan & Deci, 2000a, p. 65). For instance, you might be reading this thesis to learn what it means to be intrinsically motivated in computational creativity and videogames, for no other use but to resolve uncertainty in your understanding of the world. The other, preceding explanation has been absorbed in the concept of extrinsic motivation, characterising the engagement in an activity for a separable consequence or instrumental value. This is often accompanied by the experience of pressure and control by a third party, such as your superior asking you to write a long deliverable for the project that you are working on.

However, we can also be extrinsically motivated without such external pressure. Consider the previous example of being asked to write a long deliverable. You may feel ‘externally propelled into action’ (ibid., p. 55), and do the job cursing silently with resistance and disinterest. However, you may also accept the value of the task e.g. in keeping track of your own progress, endorse it and commit to it with a sense of volition. While you are still extrinsically motivated, your actions in the first case depend on an externally imposed, and in the latter on an internal, self-determined goal. This distinction between an external and internal perceived locus of causality (DeCharms, 1968) is thus not the same as between extrinsic and intrinsic motivations: if we talk about external causes, we always talk about extrinsic motivation. However, the presence of an internal locus of causality does not allow us to identify IM – only the other direction is true: when an organism is intrinsically motivated, they are always driven by internal causes. If you submit a deliverable to keep track of your own progress, your cause of action is internal, yet your activity is a means to an end and must be considered extrinsically motivated. But if you are intrinsically motivated, e.g.
you read this chapter out of curiosity, your actions must per definition not serve a separate outcome. Consequently, the cause of action must be internal. Ryan and Deci (2000a) hypothesise that external rewards can be internalised to different degrees, i.e. they propose a continuum between external and internal extrinsic motivation.

While these distinctions help us to understand the difference between intrinsic and extrinsic motivation to some extent, they leave us in the dark with respect to the nature of the internal causes of intrinsically motivated behaviour. Psychological theories of IM add some clarity by setting themselves apart from traditional theories of motivation and hypothesising different mechanisms that may underlie IM specifically. Crucially, these theories have inspired many computational models of IM, as shown in Sec. 2.2.4. In the next section, we survey the most important theories.

### 2.1.2 Psychological Theories of Intrinsic Motivation

How is IM functionally rooted in an organism? And what causes the selection of a specific action, in the absence of an instrumental value? The means of psychology to answer these questions are currently mostly limited to participants’ self-reports, and ‘free choice’ experiments. In the latter, most popular form to operationalise IM (ibid.), participants are asked to perform a task. After a while, they are told that they have completed the exercise, and are left with the task and a number of distractor activities. The more they keep engaging with this task in the absence of instructions and despite the distractors, the more they are deemed intrinsically motivated. Such behavioural observation and self-reports can only shed little light on the functional mechanisms of IM, and neuroscientific methods to identify and provide ‘greater resolution on the [underlying] affective and cognitive processes’ (Di Domenico & Ryan, 2017, p. 11) are only beginning to be embraced.

The landscape of existing psychological theories is twofold: some propose candidate mechanisms to explain behaviours that are typically associated with IM, while others focus on contextual factors that have been observed to support or undermine such behaviour. There is substantial overlap between these approaches, in that preferences for situational factors could be directly linked to selecting actions that lead to the same or similar situations. In what follows, we only outline a subset of theories that have inspired computational models, informed creativity and game studies, or feature in computational creativity or videogame AI. A more general, yet detailed overview of theories is provided by Di Domenico and Ryan (ibid.).

In his drive theory, Hull (1943) has rooted traditional motivation in the satisfaction of basic biological needs. A drive reduces a temporary physiological deficit to reach an equilibrium. For instance, organisms act to take in nutrients in order to maintain a certain blood sugar level. At first, researchers have tried to explain the early observations of intrinsically motivated behaviour by means of drive theory: Harlow (1950) for instance proposes the existence of a specialised drive to manipulate, and Montgomery (1954) suggests a drive to explore (ibid.). However, this extension of traditional drive theory was criticised by White (1959) for its inability to explain e.g. why
organisms engage in the exploration of an unknown environment when there is no immediate physiological deficit, or when such exploration might be detrimental to maintaining an equilibrium on essential variables.

Festinger’s (1962) cognitive dissonance theory alleviates some of the issues of drive theory by linking action-selection to the reduction of incompatibilities between perceived stimuli and internal cognitive structures such as beliefs: an organism might act to remain in a familiar environment, resulting in perceptions that confirm their belief in the state of affairs. Much later, Kagan (1972) proposed the reduction of uncertainty, in the form of an incompatibility between cognitive structures, as primary motivational mechanism. However, neither theory explains why organisms engage in e.g. activities that might lead to an increase, rather than reduction of such incompatibilities.

Hunt (1965) resolves this conflict by suggesting that children and adults prefer to trigger stimuli that achieve optimal incongruity, marking a sweet spot in the discrepancy between perceived and expected levels of stimuli. This idea is also present in Berlyne’s (1960) model of curiosity, suggesting that we prefer situations that are neither too familiar nor too novel.

A last group of theories considers effectance\(^1\) as IM, and competence, personal causation as well as self-determination as motivating factors. According to the theory of effectance motivation, proposed by White (1959) and extended by Harter (1978), humans and animals are driven by effectance to increase competence as the ability to impose an effect on the environment. The theory explains exploration as a means to develop and achieve mastery in skills. DeCharms’s (1968) theory of personal causation is closely related in that it proposes a universal propensity of individuals to ‘experience themselves as causal agents, that is, to experience their own actions as having an internal perceived locus of causality’ (Di Domenico & Ryan, 2017, p. 3). Ryan and Deci (2000b) have adopted White and De Charm’s work on effectance and personal causation for the concepts of competence and autonomy, featuring prominently as basic psychological needs in their self-determination theory. Here, perceived competence relates to effectance, and to the sense of growing mastery in activities that are ‘optimally challenging and that further develop one’s capacities’ (Di Domenico & Ryan, 2017, p. 3). Perceived autonomy in contrast relates to the experience of volition as well as choice, control or freedom of ‘either the means or ends of action’ (Ryan, Rigby & Przybylski, 2006, p. 349).

Cognitive evaluation theory (Deci & Ryan, 1985) is a part of self-determination theory with strong empirical support, demonstrating that situations which afford high perceived autonomy and competence are more likely to trigger the explorative and challenge-seeking behaviour associated with IM.

Csikszentmihalyi’s (1990) concept of flow neither considers functional mechanisms nor contextual factors of IM, but focusses on its phenomenology. It is closely related to self-determination theory, in that it describes a state of absorption and non self-conscious enjoyment of an activity, facilitated by optimal challenge and competence. It however does not recognise autonomy as essential to achieving a state of flow.

---

1 The term effectance is originally coined by White (1959, p. 329) together with the notion of efficacy as the experience produced through effectance.
2.2 Computational Models of Intrinsic Motivation

A motivational mechanism constitutes a core component of any artificial agent. Although IM represents only one type of motivation in biological systems, it has been of crucial interest for AI researchers and keeps gaining momentum. In Sec. 2.2.1, we identify and elaborate the core incentives for AI researchers to formalise IM into computational models. Formal modelling requires a thorough definition of the modelled concept; we outline the efforts of AI researchers to refine the psychological understanding of IM in Sec. 2.2.2, thus establishing the basis for a mutually informative relationship with psychology. As the core result of this chapter, we eventually synthesise their findings and our stance into a working definition of IM models in Sec. 2.2.3. We probe our working definition in the final Sec. 2.2.4, based on well-acknowledged models of IM, negative examples, and debatable cases. This working definition allows us to delineate related work in Ch. 4 and 5, and lays bare the key properties of IM to benefit computational creativity and videogame AI.

2.2.1 Incentives to Formalise Intrinsic Motivation

AI has always been strongly inspired by psychology. This influence is particularly well reflected in the first motivational mechanisms designed to guide the behaviour of artificial agents. Some robots for instance implemented Hull’s (1943) theory of drives, i.e. they acted to keep certain variables such as the energy level within fixed bounds. This allowed for high stability, but little variety and adaptivity in behaviour. An alternative option is to employ reinforcement learning (RL) (Sutton & Barto, 2018) with extrinsic reward signals, which is inspired by operant conditioning (Skinner, 1953). Here, agents use experience from interaction to learn the optimal action policy, i.e. a specification of which action to take in which situation to maximise accumulated reward signals. However, extrinsic reward signals are often sparse, hard to engineer manually, and limited to specific application domains. These challenges of extrinsic reward signals are not merely historical, but remain major constraints on (deep) RL today. When psychologists found IM as a mechanism underlying the rich,
expansive behaviour which has been considered key to human and animal development, AI researchers were keen to formalise it in computational models to overcome the limitations of existing motivational systems.

Most incentives to formalise IM have been there from the start, but are stressed with varying emphasis across times and research communities. The core incentives can be clustered into the (i) reproduction of autonomous development, (ii) the increase of task performance, (iii) generalisation, and, more recently, (iv) the compliance with and advancement of neuroscience.

One of the initial and also strongest incentives behind formalising and using models of IM in artificial agents is the (i) reproduction of human and animal ‘autonomous mental development’ (Weng et al., 2001; Oudeyer, Kaplan & Hafner, 2007). This agenda is strongly inspired by developmental psychology and observations such as motor babbling or play in infants as examples of the active, progressive and incremental learning of universally applicable skills and knowledge in the absence of an explicit task. Rather than hard-coding artificial agents, e.g. robots (Lungarella et al., 2003), with a specification of a well-understood task, they are equipped with a ‘developmental program’ which permits them to learn task-independent skills and knowledge on-line and in an open-ended way. Here, ‘knowledge acquisition’ usually refers to an agent’s improvement of its world model, and ‘skill or competence development’ denotes the learning of complex sequences of atomic actions, potentially in a hierarchical way. Just as parents help children, researchers can interact with the agents to aid learning, yet the ultimate goal is for them to exist autonomously, i.e. without human supervision (cf. Weng et al., 2001).

While employing models of IM for autonomous development can yield a wide range of fascinating behaviours (cf. Der & Martius, 2012), artificial agents usually serve the realisation of human-imposed tasks described in terms of extrinsic rewards. Yet, as already noted, such extrinsic reward signals are often sparse and hard to engineer manually. It has thus been a goal early on (Schmidhuber, 1991a; Singh, Barto & Chentanez, 2005) to complement or replace extrinsic with intrinsic reward signals in learning to (ii) increase task performance. We observe a strong surge in related research in recent years, showing that intrinsic rewards (IRs) can yield superior task performance in combination with sparse extrinsic reward, but also good performance when extrinsic reward is omitted from learning entirely. In other words, an agent which is only intrinsically motivated can yet solve tasks that matter to people. This is possible because IRs can align with the extrinsic reward implicit in such a task. We consider this the most important property of IR in this thesis. Many recent achievements centre on general game-playing (e.g. Bellemare et al., 2016; Pathak et al., 2017b; Burda, Edwards, Pathak et al., 2019), and we discuss both the underlying concept of task alignment and these advancements in more detail in Ch. 5.

A related incentive is given by the observation in behavioural psychology that knowledge and skills acquired by intrinsically motivated infants might have no immediate instrumental value when learnt, but turn out to be essential in solving unforeseen tasks. These skills thus (iii) generalise well, i.e. they prove useful in a wide range of yet unspecified applications over extended periods of time. In contrast, extrinsic rewards usually depend on a specific domain and have to be manually adapted for the agent to perform
well in a different domain. Generalisation represents one of the biggest challenges in designing artificial general intelligence (Pennachin & Goertzel, 2007), and IM holds the promise to alleviate it substantially. Crucially, task performance does not have to be strictly preceded by skill and knowledge development in a progressive fashion; Barto (2013) suggests an iterative approach, proposing that even if there is no specific task to solve right now, systems equipped with IM could remain active and ‘build the competencies needed for when they are called to action’ (ibid., p. 43).

The final core incentive for modelling IM is to increase (iv) compliance with and support new findings in neuroscience. Neuroscientists have previously found correlations between extrinsic rewards and dopamine production in the human brain and observed similarities between neural activity and the behaviour of traditional RL algorithms (Schultz, 1998). Crucially, recent studies have found increased activity in dopamine production also for intrinsically motivated exploration and mastery (cf. Di Domenico & Ryan, 2017), and have highlighted how other brain activity correlates with the learning of novel actions and the memorisation of novel information (cf. Baldassarre & Mirolli, 2013). These studies may inform future models of IM, and vice versa represent an opportunity to computationally reproduce neuroscientific findings and thus advance our understanding of IM in humans and animals.

We have highlighted the core incentives for AI researchers to formalise IM into computational models. However, we must yet identify the formal characteristics of such models that give rise to these properties, and that distinguish models of IM from other motivational mechanisms. This is necessary to understand how models of IM relate to their psychological counterpart, how they could benefit our application domains of computational creativity and videogame AI, and which characteristics models in related work and the new models developed in this thesis must fulfil to be deemed IM. In the next section, we survey existing work on defining models of IM through the lens of AI to eventually arrive at a refined working definition in Sec. 2.2.3.

### 2.2.2 Revisiting Intrinsic Motivation: Towards a Refined Definition

AI researchers interested in formalising IM have deemed the psychological concept underspecified (Oudeyer & Kaplan, 2007), and consequently revisited it with a different methodological tool-kit to produce new insights. One of our ultimate goals to benefit AI is a formal definition of IM, towards which we have made substantial progress in recent joint work (Biehl et al., 2018). Crucially, a refined definition would not only benefit AI, but could be used by researchers across the disciplines to delineate IM against other forms of motivation, and to identify underexplored areas of inquiry. Our working definition still has some shortcomings, which we elaborate on in Sec. 2.2.3. In line with Baldassarre and Mirolli (2013), we thus consider both an informal and formal definition of IM subject to ongoing work. In this section, we survey the existing research that informs our working definition of specifically models of IM in the following Sec. 2.2.3.

In order to be motivated, an agent has to choose actions based on the reward signals they have triggered in the past or are expected to yield in
the future. Our psychological account of IM in Sec. 2.1 suggests that it is not the action-selection mechanism, but the nature of the reward associated with individual actions that warrants a distinction between intrinsic and extrinsic motivation. For our working definition, it is thus less important how an agent chooses between different actions with associated (expected) rewards; we primarily focus on how rewards are assigned to actions in the first place.

To investigate IM more formally, we first have to clarify the distinction between rewards and reward signals. In psychology, rewards correspond to objects and events that attract or repel us. Reward signals in contrast are produced in response to rewards by the brain’s reward system (Barto, 2013). If not specified otherwise, we abbreviate ‘reward signals’ with ‘rewards’, which is consistent with the AI literature. We can then distinguish between internal and external rewards in terms of the location of the mechanism that generates the reward (cf. Sec. 2.1.1 and Oudeyer and Kaplan (2007)). If we are externally motivated, our actions are caused by a reward provided by some other party. But, as elaborated earlier in Sec. 2.1.1, internal rewards can yet be extrinsic. So how can we distinguish between extrinsic and intrinsic rewards? A definition of IR must address both which components of an agent’s internal or external environment are permitted to shape reward, and how.

For this endeavour, the psychological approach of discriminating IM against acting for a separable consequence or an instrumental value turns out to be particularly problematic. This is because from an evolutionary perspective, skills and knowledge acquired by means of IM must necessarily benefit an agent in successfully performing future tasks that impact survival, to justify the resource spending in their acquisition. As a very basic example, play can contribute to an animal’s development of efficient sensorimotor mappings and consequently help when escaping predators, thus increasing their fitness. Without such a handle for evolutionary forces, there would be no explanation why organisms are equipped with IM. Note that psychologists are well aware of such separable consequences when stressing the importance of IM for human and animal development. Yet, to consider a specific behaviour intrinsically motivated, they require e.g. a human participant in ‘free choice’ experiments to be either unaware of these consequences, or to not act for them (cf. Sec. 2.1.2). In contrast to the resulting psychological understatement of such instrumental outcomes, AI researchers have emphasised their presence and differentiated between the consequences of intrinsically motivated behaviour. This led to an ongoing debate on whether extrinsic and intrinsic motivations can still be considered distinct categories, or rather reside on a continuum between short- and long-term consequences, respectively (cf. Barto (2013) vs. Baldassarre (2011)). For our objective, it only matters that both sides agree that IM can indeed yield long-term consequences.

There is a separate debate whether the ultimate purpose of IM in development, realised through these instrumental outcomes, is model or skill acquisition (cf. Schmidhuber (2010) vs. Singh, Barto and Chentanez (2005)). This difference is mirrored in the distinction between knowledge-based and competence-based motivations (Mirolli & Baldassarre, 2013), representing the
key components of Oudeyer and Kaplan’s (2007) well-established typology of IM. Knowledge-based IMs depend on perceived stimuli and their relation to an agent’s expectations. Their main functional role is to support the learning of an accurate model of the agent and their environment dynamics. Competence-based IMs in contrast depend on an agent’s capacity to achieve self-determined goals as specific effects on the environment and thus support the acquisition and improvement of skills. Both types are well represented in psychology (Sec. 2.1.2): cognitive dissonance (Festinger, 1962; Kagan, 1972) and optimal incongruity theory (Berlyne, 1960; Hunt, 1965) can be associated with knowledge-based IMs, while theories of effectance (White, 1959; Harter, 1978), personal causation (De Charms, 2013) and self-determination (Ryan & Deci, 2000b) relate to competence-based motivations.

But these two types are not as distinct as they may first appear. On the one hand, model improvement relies on an agent’s capacity to perform effective and potentially complex actions. On the other hand, the efficient learning of skills requires at least partial models of the agent-environment dynamics. Schmidhuber (2010) argues that knowledge-based models implicitly introduce an objective on establishing better sensorimotor mappings, and thus subsume competence-based models. He not only requires any eligible theory or model of IM to yield model improvements, but also assumes that such improvements are the only objective in skill development. Barto (2013) in contrast highlights an agent’s competence as ultimate bottleneck on their evolutionary fitness, and thus defers model improvement as a mere facilitator for skill development. This ‘competence view’ is articulated softer than Schmidhuber’s (2010) in that it acknowledges the importance of model improvement, yet emphasises skill development as the ultimate aim of IM. These two positions inform our fourth diagnostic of IM models in Sec. 2.2.3: the capacity for open-ended development.

The insight that IM yields long-term consequences – no matter the ultimate purpose – represents a first step towards refining the concept. For behaviour to be valuable across a long time horizon, IR must be independent of factors that are specific to a certain situation. This may explain the ignorance of researchers and participants with respect to the immediate consequences of intrinsically motivated behaviour in psychological experiments. To account for this independence, Barto (2013) argues that IRs are more likely to be shaped by components of an agent’s *internal environment*:

> ‘Novelty, surprise, incongruity, and other features that have been hypothesized to underlie intrinsic motivation all depend on what the agent has already learned and experienced, that is, on its memories, beliefs, and internal knowledge state, all of which are components of the state of the organism’s internal environment.’

---

3 The original typology entails morphological models as a third type for which actions are chosen based on the properties of sensorimotor values at different timesteps and irrespective of the agent’s current predictive model quality or skills. For instance, reward could depend on the ‘high short-term correlation between a maximally large number of sensorimotor channels’ (Oudeyer & Kaplan, 2007, p. 10). We have omitted this type for its presently low popularity and because its consequences on the agent and their interaction with the environment are not clear. The original typology also contains further subtypes, but Schmidhuber (2010) shows that their boundaries are blurry and we consequently omitted them as well.
(...) the internal environment may play a larger – or at least, a different – role in generating reward signals associated with intrinsic motivation’ (Barto, 2013, p. 36).

Barto (ibid.) herewith provides a compelling argument for the dominance of agent-internal components in shaping IR. But we cannot deny that ultimately, the external environment can affect IR indirectly: it enters the agent’s internal environment through their sensors, and can then influence internal components such as beliefs and memory. While an embodied agent cannot access the external environment directly, it yet appears reasonable to assume that their perception of and beliefs about the environment could contribute to and shape IR. This raises the question of how IR can integrate such indirect influences of the external environment while being invariant to specific niches or situations, thus preserving the long-term value of the resulting behaviour.

An explanation to this is implicitly given by Oudeyer and Kaplan (2008), who characterise IM inductively based on a survey of existing models and psychological theories. Key to their account of IM is the concept of collative properties, which Berlyne (1965) has coined in his studies of intrinsically motivated exploration. Berlyne has observed that stimuli from the external environment that trigger exploratory behaviour score particularly high on properties such as novelty, surprisingness or incongruity. He moreover noticed that these properties are of a special ‘collative’ kind, in that they ‘all depend on the collation or comparison of information from different stimuli elements’ (ibid., p. 246), and can thus be formalised with information-theoretical measures based on an agent’s subjective uncertainties in a given environment. Oudeyer and Kaplan (2008, p. 95) adopt this concept in their characterisation of IM:

‘An activity or an experienced situation, be it physical or imaginary, is intrinsically motivating for an autonomous entity if its interest depends primarily on the collation or comparison of information from different stimuli and independently of their semantics, whether they be physical or imaginary stimuli (...) perceived in the present or in the past (...) or stimuli that are simultaneously present in different parts of one stimulus field.’

(Oudeyer & Kaplan, 2008).

Oudeyer and Kaplan thus allow us to complement Barto’s account of what shapes IR signals by an account of how they are formed.

While the debate is ongoing, we find that AI researchers, by scrutinising the psychological concept of IM from a computational but also biological perspective, contribute to a better understanding across the disciplines. They assert that IM must indeed have long-term consequences, and put forward functional theories on how IR is shaped and by what. Next, we synthesise these insights and our stance in a working definition of IM models.

2.2.3 Models of Intrinsic Motivation: A Working Definition

In this section, we establish a working definition of IM models that complements the psychological account. Our definition is informed by the insights introduced in the previous section, and by a formal definition jointly put
forward in (Biehl et al., 2018). Despite these formal advances, our working definition is decidedly informal. We briefly justify this choice. The formal treatment in (ibid.) relies on a predictive formulation of IM, i.e. it assumes that actions are chosen with respect to the future expected reward they are predicted to yield. This mechanism necessitates an agent-internal generative model to predict the future (latent) state and the dynamics of the environment and of agent-internal components. We believe that most existing models of IM either already comply with this predictive formulation or can be reformulated while retaining their characteristic emergent behaviour. However, our formal definition is complex and might not allow for a quick assessment of whether a specific model adheres to it. In order to match existing models in related work more easily and without reformulating them in terms of predictions, we sacrifice the formal approach in favour of a more intuitive informal treatment. That said, we still introduce empowerment maximisation (EM) as our core model of IM in a formally rigorous manner in Ch. 3.

We have earlier identified the nature of reward as the discriminator between intrinsic and extrinsic motivation and we consequently base our definition of IM models on the properties of an agent’s action-value function. The function maps a sequence of actions of arbitrary length to a scalar reward, based on how these actions have or are expected to influence the components that contribute to reward\(^4\). The more of the following properties the function fulfils, the more we consider its output an IR\(^5\):

- **Agent-centricity**: Intrinsic reward is computed from the perspective of an embodied agent with potentially limited means to perceive and affect their environment. The action-value function must thus only use agent-internal components, i.e. sensors, actuators and an agent’s internal state, the latter which can be further unravelled into memory, beliefs, etc., depending on how the agent is modelled.

- **Freedom of semantics**: Intrinsic reward is shaped by the distribution of values of, and by the relationships between, agent-internal components. It is invariant to a component’s specific values and permutations thereof, as well as to the meaning of the component itself. The action-value function must thus be free of semantics on both levels.

- **Embodiment universality**: Intrinsic reward should be both computable for, and yet remain sensitive to any possible agent embodiment. The action-value function must thus produce different rewards for any possible coupling between an arbitrary number of sensors and actuators.

- **Open-ended development**: The action-value function, given a specific action-selection mechanism, must drive an agent’s behaviour towards ongoing model and skill improvement for a large majority of possible embodiments and situations, leveraging all available degrees of freedom in the agent and their environment.

---

\(^4\) This is similar to the action-value function in RL, but we assume that a value can be provided for an action sequence of arbitrary length, rather than for a single action and a policy.

\(^5\) One might still wonder whether a particular intrinsic action-value function could be considered a better fit than another for a specific agent. We elsewhere (Guckelsberger & Salge, 2016) discuss the requirements for an arguably more natural choice of such a function.
The first two properties follow from the previous observation (Sec. 2.2.2) that IR is decoupled from any short-term consequences on specific situations. With the latter two properties, we want to account for the presence of the characteristic, generalising behaviours in a large range of organisms. The property of open-ended development complements embodiment universality in that it requires different embodiments to not only yield diverse, but also sensible rewards. For the operationalisation of sensibility, we embrace both the knowledge and competence views without giving either priority. Note that, rather than taking the theoretical considerations from Sec. 2.2.2 as a starting point, we can also pragmatically justify the selection of these properties as facilitators of the incentives to embrace the IM concept in AI (Sec. 2.2.1).

We ask the reader to understand these properties as diagnostics, and not as necessary or sufficient criteria: the more of them apply, the more we are comfortable to call the output of the action-value function an intrinsic reward (IR), and a model which uses these rewards for action selection an intrinsic motivation. We have chosen this diagnostics formulation for a number of reasons. Firstly, based on the insight that IMs also yield separate outcomes (Sec. 2.2.2), we favour a gradual rather than a binary distinction between these two types of motivation. Secondly, we would like to emphasise the individual properties and their implications for our applications over a clear-cut distinction. Thirdly, we could challenge every diagnostic with pathological cases that can rarely be found in nature. For instance, we could define embodiments that exploit a given value function to yield behaviour opposed to open-ended development. Finally, we do not believe that a binary and yet universally accepted definition of IM models is presently feasible, given their inspiration by and relationship to the original but underspecified concept in psychology. Note though that despite this formulation as diagnostics, we consider some properties as more important than others, with the order above roughly reflecting high to low priority.

Some of these diagnostics are harder to assess than others. We can check agent-centricity by considering whether the variables used in a specific action-value function correspond to agent-internal components only. One appealing intuition for the freedom of semantics is that the action-value function should be ‘rewiring agnostic’: IR must be invariant to either a virtual or physical rewiring between the agent-internal components and the action-value function. In other words, all internal components of the same kind, e.g. different sensors, must be treated alike. Given freedom of semantics on the level of components, we can follow that no semantics are used on the level of their values. More formally, we are free of semantics if we can apply a bijective transformation on the values of a specific component as the input to the action-value function without changing the calculated reward. Many existing models of IM satisfy this property by relying on information-theoretic measures (Appx. C), or alternative probabilistic options (cf. Oudeyer & Kaplan, 2008). We assess embodiment universality with respect to the concept of objective sensorimotor embodiment defined in Appx. D. For the sake of simplicity, we assume agents to have unlimited processing power and memory capacity. An action-value function must then be computable for, yet be sensitive to any possible shape of the distribution in Eq. D.3 which encodes the coupling between an agent’s actions and future perceptions. Crucially, this entails any possible sensor,
actuator and environment dynamics as determinants of the coupling. An eligible action-value function must work and produce different rewards for an underwater robot equipped with a sonar sensor and fin actuators, but also for a drone with a camera and propellers as well as any other combination of sensors, actuators and environment. In practical terms, given freedom of semantics and thus independence of specific component types, an eligible action-value function must be able to use an arbitrary amount of sensor signals in the calculation of reward, and assign it to an arbitrary amount of actions. Embodiment universality for actual models of IM can be proven analytically, and we use proof by contradiction to reject other motivational models in the next Sec. 2.2.4. Open-endedness requires IR not to flatten out until the agent has fully exhausted the potential of its embodiment for such improvement, and a qualifying model must prevent the agent from getting stuck in situations where such development is inhibited. Such situations of e.g. sensory impoverishment are often referred to by the metaphor of a ‘dark room’ (e.g. Friston, Thornton & Clark, 2012), which we adopt for our evaluations. In the opposite, the development of intrinsically motivated agents can also be inhibited by local attractors that provide insatiable reward (cf. e.g. Burda, Edwards, Storkey et al., 2019). A systematic quantification of open-ended development remains a grand challenge, with existing approaches (e.g. Merrick, 2008b) being not universally applicable. Presently, we rely on thought-experiments and empirical data on the induced long-term behaviour.

Our working definition distils the understanding of IM from a computational perspective which was formulated in response to psychology, and allows us to work out the commonalities and subtle differences between both accounts. Recalling from Sec. 2.1.1, psychology defines IM as engaging in an activity for its inherent satisfaction, rather than for a separable consequence or instrumental value (Ryan & Deci, 2000a). In a computational context, the ‘engagement in an activity’ is captured by an agent’s ongoing choice of action, based on the reward assigned to these actions by the action-value function. While psychology operationalises IM based on participants’ behaviour via self-reports or observations in ‘free choice’ experiments, we use four diagnostic properties on the action-value function. To be deemed intrinsic, both psychology and AI require reward not to be external of an agent, perceived through their sensors as a separate entity, e.g. in the form of an instruction, but to be produced within. In psychological terms, an agent must have an internal locus of causality, and we enforce this with the properties of freedom of semantics and agent-centricity, respectively. However, as noted in Sec. 2.1.1, such an internal locus of causality is not sufficient to rule out extrinsic motivation; an agent must also be denied to act for a separate, instrumental outcome. This is also warranted by freedom of semantics: without presupposing semantics, an agent cannot act towards a specific internal state articulated e.g. as a belief. We thus cannot a priori direct behaviour to the achievement of a specific outcome, similar to Harlow (1950) who could not post-hoc relate the behaviour of animals to a specific separate outcome, and consequently coined it ‘exploratory’ and ‘playful’ (Ryan & Deci, 2000a). Instead, an agent entailed with a model of IM can only act on the distribution of and relationship between e.g. sensor values as abstractions of such states, yielding a heuristic to potentially produce sensible behaviour across many situations. An agent
may well be able to predict these separate outcomes, e.g. via planning, but it is not permitted to use this knowledge in reward calculation. In contrast to psychology, the AI account of IM emphasises the importance and nature of instrumental outcomes, especially since these enable generalisation and allow intrinsically motivated behaviour to complement as well as substitute extrinsic reward for increased task performance (Sec. 2.2.1). As a subtle difference, while psychology seems to allow intrinsically motivated beings to be aware of such outcomes as long as they do not guide their behaviour, it does not strictly require them. We could thus in principle conceive of a psychological theory of IM with actions that have no instrumental value, and hence conflicts with the evolutionary argument made by AI researchers (Sec. 2.2.2). Our definition consequently explicitly asks for a model of IM to contribute to an agent’s open-ended development. A final difference concerns the relationship of psychological theories and models of IM with respect to different embodiments. Psychologists have observed similar, intrinsically motivated behaviour in both people and animals, and developed theories to explain the functional underpinnings. However, there is little evidence that different organisms actually realise the same hypothesised principle functionally, or only behave as if this was the case (cf. Sec. 3.4). In our AI definition in contrast, we require a specific model of IM to be embodiment universal.

We briefly evaluate our four diagnostics in the context of videogames. To warrant agent-centricity, the action-value function must not require access to the global game state or an objective account of the game’s dynamics; instead, it must be possible to calculate IR from the perspective of an agent in the game, such as the player. This perspective rests on the agent’s unique means to perceive and act in the game, determined by their embodiment, and their models of the game state and dynamics. To be free of semantics, the action-value function must be agnostic with respect to the meaning of game tokens as determined by the player or designer: it must not matter whether they are facing a power-up, a weapon, or in particular any kind of reward. An action-value function fulfils embodiment universality, if it can be computed for any agent that could interact with the game world, and is yet sensitive to their abilities to affect and perceive it. It should yield different rewards for e.g. a player that can walk but also for an NPC, who perceive the world in fundamentally different ways. Finally, the intrinsic value function, in combination with a specific action-selection mechanism, should allow for open-ended development by driving an agent towards enhancing their model of the game world and their abilities within. In particular, it must not render the agent stuck in set of states that is small in comparison with what they could potentially realise in interaction with the game mechanics.

Equipped with a working definition of IM models, we can resolve ambiguity in the relationship between models of IM and intrinsically motivated RL. Already in early work by Schmidhuber (1991a) as well as Singh, Barto and Chentanez (2005), intrinsic action-value functions have been used as a source of IR in RL, optionally in combination with an extrinsic reward signal. This requires us to understand the reward function in (partially observable)

---

6 In Ch. 5, we note that this agent is not necessarily restricted to characters such as the player or a non-player character (NPC), but could also be a procedural content generator.

7 This is not to be confused with Singh, Barto and Chentanez’s (2005) homonymous model.
Markov Decision Processes (cf. Appx. D) as part of an agent’s internal environment (Barto, 2013). An agent then leverages experience to learn how much reward can be expected in specific situations for a given action as the input to a reward maximising policy, to learn such a policy directly, or both. Closely related, Srivastava et al. (2019) recently proposed upside down RL to bridge the gap between supervised and reinforcement learning. Here, a state, and a command in the form of a desired return, potentially shaped by intrinsic reward and to be achieved within a specific time horizon, are inputs to a behaviour function, which is subjected to supervised training to produce the actions which are most likely to realise the command, starting from the state.

There exist further means to leverage agent experience in connection with intrinsic action-value functions, which lead to some ambiguity. For instance, an agent might use RL to acquire a predictive model as a requirement for the intrinsic action-value function (e.g. Burda, Edwards, Storkey et al., 2019), by treating the negative model error as a reward to maximise. Furthermore, some IRs are hard to compute directly but can be estimated by sampling from the experience of ‘partial rewards’ (e.g. Gregor, Rezende & Wierstra, 2017). IR functions that rely on one or multiple forms of these learning mechanisms can yield different rewards in the same situation at different times, and are consequently referred to as adaptive motivations (Oudeyer & Kaplan, 2007). Crucially though, there also exist static (ibid.) models which do not leverage learning from experience at all. Intrinsically motivated RL is thus not synonymous to, but represents a subset of models of IM.

This concludes our working definition of IM models, consisting of four diagnostic properties for an action-value function to yield IR as the basis for IM. We next validate this proposal based on well-acknowledged models of IM, as well as negative examples, and we highlight borderline cases.

### 2.2.4 Example Models of Intrinsic Motivation

We have chosen the following examples of motivational models – intrinsic, extrinsic and controversial – not to be exhaustive, but to illustrate their diversity and to provide close reference points for identifying similar models of IM in the related work in Ch. 4 and Ch. 5. We use the large diversity in well-acknowledged intrinsic and extrinsic models of motivation to demonstrate the power of our working definition, limited to two models in each category to avoid redundancy. We demonstrate that the differences between these categories are often not clear cut, based on two examples that are usually considered IM, but only qualify as borderline cases according to our definition. We describe each model functionally and hypothesise the potential emerging short- and long-term behaviour. Similar to our working definition though, we abstain from a rigorous formalisation to allow for an easier comparison with slightly different models. We maintain a chronological order to highlight influences in the development of these models, and we point out their relationship to psychological theories (cf. Sec. 2.1.2) where applicable.

Historically, psychological theories of IM focussed on exploratory behaviour, and it is thus not surprising that the first models of IM capture **curiosity**.  

---

8 These cases correspond loosely to action-value, policy gradient and actor-critic methods.
While the earliest formalisations go back to the 70s and 80s (e.g. Lenat, 1976; Scott & Markovitch, 1989), most contemporary research is based on ideas developed throughout the 90s (cf. Schmidhuber, 2010). In the arguably most well-known formalisation, an agent acts to increase the prediction error of their world model (Schmidhuber, 1991b) as IR. They consequently provoke unfamiliar perceptions, and thus engage in exploration. This approach is closely related to Festinger’s (1962) cognitive dissonance theory. We refer to this and similar formal models as artificial curiosity.

The prediction error formulation of artificial curiosity meets all properties of our working definition. Its action-value function only involves an agent’s world model and sensor inputs, i.e. it is based on agent-internal components alone and is thus agent-centric. It is free of semantics, in that the distinct sounds of a tweeting bird and a gunshot would give rise to the same IR if they were predicted with equal probability by the model, irrespective of their meaning. This freedom of semantics is also demonstrated by the model’s rewiring agnosticity, i.e. the action-value function would generate the same IR even if e.g. audio and visual sensors had been switched. Given sufficient processing power and memory capacity, we can scale a world model to predict the effect of an arbitrary number of simultaneous actions on any number of future perceptions. Artificial curiosity is thus embodiment universal. Finally, it realises open-ended development as in the majority of situations, an agent driven by artificial curiosity would keep exploring and improving its world model, while avoiding dark rooms where prediction error is zero.

The exception to these situations represents the major caveat of the prediction error maximisation approach to artificial curiosity: an agent can become attached to random sources of noise, recently denoted ‘noisy TV’s’ by Burda, Edwards, Storkey et al. (2019), as these remain unpredictable and the intrinsic reward in the respective states does not satiate. To overcome this, Schmidhuber (1991b) has early on proposed an alternative motivation for which an agent chooses actions that maximise their expected learning progress.

In a popular version, this is quantified as the difference in the prediction error of consecutive future sensor states, provided by an agent’s world model (Schmidhuber, 2010). Oudeyer, Kaplan and Hafner (2007) extend the original formulation by splitting the state space into regions, which increases robustness. Similar to Schmidhuber’s (1991) version, their ‘intelligent adaptive curiosity’ agent stays away from situations that are either too predictable or unpredictable. This family of models is thus closely related to Hunt’s (1965) optimal incongruity theory and Berlyne’s (1960) adaptation to curiosity. In these models, learning progress as the difference in prediction error is either calculated explicitly (Oudeyer, Kaplan & Hafner, 2007) or implicitly via separate predictive models (cf. Schmidhuber, 2010). In either case, they assume the existence of a reliable distance measure on sensor space.

Knowledge-seeking (Storck, Hochreiter & Schmidhuber, 1995; Orseau, 2014) is a more principled approach to modelling learning progress which is free of such assumptions, but usually intractable. Learning progress is quantified as an agent’s information gain, the amount of additional information about the latent environment state and the environmental dynamics acquired through a particular sequence of future sensor states triggered by the agent’s actions. Each possible parametrisation of the model can be considered a different
hypothesis of how the agent can affect their future perceptions, mediated by the environment. Maximising information gain then allows an agent to select those actions which optimally probe and limit their hypotheses about the environment. It also likely yields a perfect model over the long term.

The previously described models of learning progress motivation also score on all diagnostics of our working definition. Consider knowledge-seeking as an example: The underlying information gain is given by the mutual information (Appx. C) between an agent’s (expected) sensor perceptions on one side, and their (expected) change in beliefs about both the latent environment state and the environment dynamics on the other side. It is thus both agent-centric and free of semantics. The latter is illustrated by the fact that applying a bijective transformation on both sensor and belief states would not change the mutual information. To get different information gain rewards, an agent’s embodiment must only afford that different actions yield different sensory futures, and that these give rise to different internal belief states. For several physical or virtual sensors and actuators, we can encode these states as vector-valued random variables and the model is thus embodiment universal. Finally, knowledge-seeking agents realise open-ended development; they do not get hooked on noise and avoid dark rooms as the corresponding perceptions would offer no information gain about the environment.

Both, the prediction error and learning progress formulations of artificial curiosity, have traditionally suffered from scaling badly to large state and action spaces. However, with the availability of better function approximation based on (deep) neural networks paired with RL, these models have again gained strong momentum in recent years (e.g. Bellemare et al., 2016; Pathak et al., 2017b; Burda, Edwards, Pathak et al., 2019).

The previous models are knowledge-based (cf. Sec. 2.2.2), in that rewards depend exclusively on the prediction quality and improvement of an agent’s model. Specific future perceptions associated with different sensorimotor mappings only feature as a means to achieving more favourable model updates. If we fixed the model parameters – i.e. the model could not incorporate new sensorimotor evidence – exploration would likely fail. Predictive information maximisation (Ay et al., 2008; Ay et al., 2012) in contrast yields exploratory behaviour but sits half-way between knowledge-based and competence-based models. This is because IR here not only depends on the quality of an agent’s model, but also directly on the possible sensory futures their actions can produce. A predictive information maximising agent would yield an exploration strategy even if their model parameters were fixed (cf. Ay et al., 2008).

Introduced as a natural complexity measure for time series (Bialek, Nemenman & Tishby, 2001), predictive information can be calculated on an agent’s sensor state space to measure how much information a preceding sequence of perceptions holds about a consecutive sequence. By adjusting the parameters of an agent’s policy to maximise this quantity, an agent realises a trade-off between causing a rich, diverse sensory future while keeping it as predictable as possible from past perceptions. Simulation studies (Ay et al., 2008) show that this maximum coincides with the agent’s ‘effective bifurcation point’,

---

9 More precisely, artificial curiosity agents would likely get stuck in states that yield the momentarily highest prediction error. Learning progress motivations in contrast would flatten, potentially reducing exploration to a random walk, depending on the actual implementation.
a dynamical systems concept describing the critical region in policy parameter space where different stable behaviours can be distinguished, but are yet sensitive to the environment and switched upon perturbations (Der & Martius, 2012, pp. 41–44). An agent would try to cause the most complex possible sensory process, and thus engage in exploration. At the same time, they would switch between stable behavioural modes e.g. when running into obstacles. They would avoid random noise sources as these cannot be predicted from past sensor values, and they would not get stuck in a dark room as it would not allow to distinguish different sensor states. If predictive information is measured on arbitrarily long sensor sequences (cf. Martius, Der & Ay, 2013), agents are expected to establish an ergodic process where every possible sensor value is created in a complex sequence which is then looped to maintain predictability$. For short sensor sequences, we get ‘playful’ behaviour (Der & Martius, 2012) in which agents exercise different behavioural modalities in response to the environment. This model has been inspired by ‘homeokinesis’ (ibid.), a dynamical systems theory developed in opposition to the homeostasis concept in Hull’s (1943) traditional drive theory.

Based on the AI literature in particular, we may get the impression that IM is synonymous with exploration; however, taking into account its conceptualisation in psychology and AI, intrinsically motivated behaviour is much broader. The free energy principle (Friston, 2010) demonstrates this nicely, as it has attracted strong attention both in psychology and neuroscience, as well as in AI. It has been proposed as a unified theory explaining a wide range of cognitive processes at different levels of complexity, from the single cell to the human brain. The free energy principle has been motivated by the observation that biological systems must limit the range of perceived sensor states to maintain self-organisation (ibid.). In its basic formulation, the free energy reward corresponds to an agent’s uncertainty about their future sensor states, given the (latent) environment states they can trigger with their actions (cf. Friston, Parr & de Vries, 2017, supporting material Eq. A.2). However, in contrast to the previous exploration models, an agent realising this principle is assumed to minimise free energy and thus decrease rather than increase the uncertainty of future sensor perceptions. It chooses actions which are expected to lead to (latent) environment states perceived via precise, unambiguous sensor values. The free energy principle therefore bears similarities with Festinger’s (1962) cognitive dissonance theory and Kagan’s (1972) adaptation as reduction of uncertainty.

A free energy maximising agent will avoid random noise, but may get stuck in a dark room as the latter only yields a single sensor state. This can be partly alleviated by extending the action-value function by a term which is identical to the information gain in knowledge seeking (cf. Friston et al., 2017). Free energy can be minimised by both performing actions to yield more precise perceptions, or by changing the model that encodes the agent’s beliefs about future perceptions given (latent) environment states. This joint optimisation is called active inference$^\text{(Friston, 2010)}$. This supports our earlier observation that the boundaries between knowledge- and competence-based IM are fluid.

---

10 Based on personal exchange with Martin Biehl, referring to a discussion with Georg Martius.
11 In (Biehl et al., 2018), we generalise active inference to employ other IRs beyond free energy.
We soon argue that the free energy principle represents a controversial example of IM. But first, we probe our definition on negative examples, i.e. models which are generally accepted to not resemble IM. We return to the theory of drives (Hull, 1943) which served as the primary explanation for motivation before the conceptual divide (cf. Sec. 2.1.2). A hunger drive, i.e. a motivation to reduce a physiological food deficit and reach an equilibrium, is usually considered an internal but extrinsic motivation. Albeit being agent-centric, an implementation of a hunger drive would score low on the other diagnostics of IM models because reward would have to be formulated relative to an agent’s specific digestive system; we would have to distinguish between different nutrients in terms of their quality or toxicity, and thus would not be free of semantics and as an implication, not embodiment universal. This also implies that drives are not rewiring agnostic, i.e. we cannot arbitrarily exchange e.g. different sensory inputs. Furthermore, while a hunger drive might allow for short-term model and skill development, it does not warrant open-ended development: as soon as an agent had found a niche that supplies enough nutrients, they would reside there and stop improving their skills and knowledge. Further development is thus inhibited through the consummatory climax inherent to drives.

More closely related to our application domain, most motivations used in general game-playing would not qualify as intrinsic. As a concrete model, consider Mnih et al.’s (2015) well known deep q-network RL agent. Here, the action-value function uses raw pixel data of the game, and a reward which represents the change in game score. The network learns to associate different perception-action tuples with expected accumulated reward in RL. Since the score has been defined by the game’s designer, and is communicated as a score difference to the agent by the model engineers, it corresponds to an external and extrinsic reward (cf. Sec. 2.1.1) and is subject to the sparsity and engineering challenges detailed in Sec. 2.2.1. This model does not diagnose as IM at all; most crucially, it relies on an agent-external component and is thus not agent-centric. But even if we considered the reward signal to be fed to the agent through a sensor, we would have to distinguish between higher and lower sensor values and would thus not be free of semantics. Since the reward sensor would be treated specially, we would imply semantics on the level of components and not be rewiring agnostic. The model would lack embodiment universality, in that it would only be sensitive to sensors that perceive the reward it is tuned to, and to environments that afford the reward in question. Since this would only be the case for very few environments, the model would not score on open-ended development either.

We finish with two more controversial examples that score on some but not all diagnostics of our definition. Our goal is to highlight the divergent views on what should be considered IM, and the need for further research to resolve these discrepancies. Singh, Barto and Chentanez’s (2005) intrinsically motivated reinforcement learning (IMRL)12 has been emphasised as the first competence-based model of IM (Mirolli & Baldassarre, 2013). It provides a

---

12 We use the abbreviation IMRL to discriminate this specific motivational model from the more general technique of employing IR in RL.
2.2 Computational Models of Intrinsic Motivation

RL agent with rewards to learn complex skills which can trigger salient events, such as opening a door. Skills are formalised as options (Sutton, Precup & Singh, 1999), i.e. actions sequences that terminate in a certain state set. For some termination states, a salient event is triggered and the agent receives a reward proportional to how unlikely they expected to reach these states. The agent hence keeps learning an option until performing it reliably well, and then moves on to gradually chain options together into complex sequences to trigger more remote events. In spite of its name, we question the intrinsic nature of the model, as the salient events that trigger IRs are hard-coded.

A similar issue applies to the free energy principle. Most formalisations (e.g. Friston et al., 2015) go beyond the ‘core formalism’ distinguished earlier by minimising an additional term which quantifies the distance between the predicted sensor states following the agent’s actions and a desired distribution over future perceptions. An agent with this extended model will act to approach this distribution. Friston, Thornton and Clark (2012) argue that it is this expectation over a desired sensory future that eventually saves a free energy minimising agent from the dark room problem. The literature is inconclusive with respect to whether this formulation is still considered an IM. It poses an interesting challenge, which we briefly elaborate here.

In contrast to the score in Mnih et al.’s (2015) general game-playing agent, Singh, Barto and Chentanez’s (2005) salient events and the sensor target distribution in free energy are not externally imposed, but internal to the agent right from the start of an experiment. We thus cannot reject the models on the basis of agent-centricity. However, we find it problematic to consider the corresponding models intrinsic, as the reward computation includes priors that are invariant with respect to the state of other internal components. Friston, Thornton and Clark (2012) argue that the prior in free energy is defined by evolution, a point which could be equally made for the salient events in Singh, Barto and Chentanez’s (2005) model. But how is such a prior different from any other hyperparameter provided to the model at the onset? Could a motivational model as a whole not be considered an evolutionary prior on an agent’s behaviour? These questions highlight the need for further research. At this point, we consider both motivations as borderline models of IM, because their priors limit the optional property of open-ended development. Once e.g. Singh, Barto and Chentanez’s (2005) agent has exhausted all pre-defined salient events, they might continue improving existing skills, but further skill development will cease.

In this chapter, we have introduced the concept of IM both through the lens of psychology and AI. We have discussed its origins, and elaborated on the distinction between internal vs. external, and intrinsic vs. extrinsic, motivations. We have summarised the most popular theories of IM in psychology, and explained the various incentives to formalise it in computational models. We have reviewed the efforts of AI researchers to disambiguate the psychological concept and, based on this ongoing research, devised our working definition of IM. This definition has been evaluated on a representative set of well acknowledged reference models of IM, negative and some controversial examples that highlight both its power and limitations. We believe that our definition is sufficiently precise to allow for the identification of IM models in
our related work in Ch. 4 and Ch. 5, a task which is further supported by the introduced reference models. Crucially, empowerment maximisation (EM) as the central model of IM to be used in this thesis has not been addressed yet because it requires a comprehensive and formally rigorous introduction. This is the subject of the next chapter.
Empowerment maximisation (EM) is the central model of IM to be investigated, extended and applied in this thesis. Originally introduced by Klyubin, Polani and Nehaniv in 2005, it is based on empowerment as IR reward, an information-theoretic quantity which measures an agent’s influence on their environment. The goal of this section is to introduce EM in an intuitive but also formally rigorous way, and to sketch related research as context for our novel contributions in computational creativity and videogame AI.

In Sec. 3.1, we motivate empowerment as IR and EM as a model of IM via empirical observations and theoretical considerations in psychology, biology and physics. We then define empowerment and the corresponding motivational model both informally and formally in Sec. 3.2. We dedicate Sec. 3.3 to discussing the properties of empowerment and to evaluating EM against our working definition of IM models. To support the reader’s intuitions, we use examples that showcase the empowerment quantity and the behaviour of an empowerment maximising agent. In Sec. 3.4, we finally differentiate the related work landscape, thus establishing reference points for the contributions of this thesis to empowerment research and vice versa.

Following its inception in 2005 (Klyubin, Polani & Nehaniv, 2005a, 2005b), empowerment and EM have been conceptually penetrated to increasing depth, and the original formalism has been refined and generalised several times. In addition to providing the aforementioned overview, this chapter also contributes a rigorous and up-to-date formalisation of EM informed by recent joint research in (Biehl et al., 2018). Most notably, we resolve some formal ambiguity in earlier work with respect to the modelling of an agent’s subjective and limited perspective on their environment. More specifically, we make an explicit formal distinction between the agent’s embodiment in an objective, physical world and the calculation of empowerment as IR based on their beliefs about that world. Although we simplify the empowerment calculation to a certain extent in our later experiments, we introduce it in a more general fashion as basis for a discussion of future work. This chapter can be considered an update of the earlier EM survey by Salge, Glackin and Polani (2014b), with a focus on discrete empowerment.

3.1 MOTIVATION

Presented with the choice to be poor or wealthy, sick or healthy, and imprisoned or free, most people would not deliberate for long (ibid.). We would usually express a strong preference for increasing our wealth, health and freedom, although our current plans or goals do not require it, and the impact of these activities on our future lives remain nebulous to us. But what is the common theme underlying these preferences? And what mechanism allows us to respond to each of these choices without the effort or possibility of contemplating their long-term consequences?
Being wealthy, healthy and free provides us with more degrees of freedom with respect to the things that we can do, and increases the number of possible outcomes that we can achieve. These are flip sides of the same coin, expressing a stronger potential and perceivable control over our environment: in each of the preferred situations, we expect our actions to have a stronger influence on ourselves and the world that we live in. Crucially, this equips us with a high level of preparedness towards unforeseen events. Based on these situational preferences, we can identify a unifying behavioural principle to explain each of the earlier choices, a local heuristic that allows us to make decisions with potentially inconceivable, global, long-term consequences: all else being equal, increase your options and influence (Klyubin, Polani & Nehaniv, 2005a).

We can identify specialised behaviours deduced from this principle throughout the animal kingdom and at many levels of complexity: Sugar-feeding bacteria prefer locations with high sugar concentration as they bear more possibilities for further locomotion, and better chances for reproduction (Klyubin, Polani & Nehaniv, 2008). Similarly, chimpanzees try to increase the social status within their troop to have more mating choices (Klyubin, Polani & Nehaniv, 2005a). For us humans, having more money, better health and more freedom usually increases our chances for long-term survival and well-being. We have even included this heuristic in our artefacts: in the board game Reversi (Othello), players can perform best on average if they increase their mobility, i.e. the number of moves they can make in any given situation (Klyubin, Polani & Nehaniv, 2005b). A single behavioural principle to have more potential and perceivable control unifies the wide range of seemingly disparate drives towards higher sugar concentration, social status, money, health, freedom and mobility.

Psychologists have discussed the importance of control in people’s situational evaluation and behaviour already at the turn of the 20th century: in ‘The Play of Man’, Groos (1901) points at peoples’ experience of joy when being able to control sensory stimulation. More than 50 years later, White (1959) reports observations of people repeatedly engaging in behaviours that have an effect on their environment independently of a physiological need. In his theory of effectance motivation, he consequently proposes such control as driver of intrinsically motivated behaviour (cf. Sec. 2.1.2). This has been further extended by Harter (1978) and adopted in DeCharms’s (1968) theory of personal causation. Referencing White, Watson (1966) describes contingency awareness, i.e. an organism’s functional knowledge that the nature of a received stimulus is sometimes affected by the nature of executed behaviours, as key factor in the development and learning of early infants. Similar to White (1959), he suggests that a stimulus becomes associated with a reward as result of being contingent on a particular action. A contingency aware organism is sensitive to and set to learn and repeat stimulus-response patterns. Rotter (1966) proposes that individuals distinguish situations in terms of the type of perceived control they afford, given their available actions, perceptions and expectations. He situates an individual’s perceived control between the two loci of external vs. internal control. In the case of an external locus of control, a subject attributes a mismatch between their expectation and the observed outcome of an action entirely to external forces. In the case of
an internal locus, an individual explains control entirely in terms of their own actions invoking the observed outcome. Closely related to the notion of an external locus of control, Seligman (1975) argues that people learn to avoid states where they feel helpless, i.e. where their actions seem to have random or no outcomes. He proposes that being in such states can have negative consequences on well-being, and describes a behavioural tendency to avoid situations with little internal control. Oesterreich (1979) proposes a similar behavioural principle in a quantitative, stochastic framework, in which he associates chains of actions with potential follow-up perceptions via transition probabilities. He characterises states by their ‘efficiency divergence’ (German: ‘Effizienzdivergenz’), representing the likelihood of different actions leading to distinct follow-up states. Oesterreich proposes that people should act towards states with maximum efficiency divergence. His focus is not on well-being, but he suggests that such behaviour could benefit explicit, potentially long-term goals. The seeking of potential control also resonates with the notions of competence and autonomy in Ryan and Deci’s (2000) self-determination theory. Related, the cyberneticist von Foerster (1984/2003a, 1973/2003b) proposes a simple behavioural imperative: ‘I shall act always so as to increase the total number of choices’ (von Foerster, 1984/2003a, p. 282).

These psychological theories consider control from an individual’s subjective perspective on the world. They incorporate a notion of perception which has been explicitly put forward by Gibson (1979), arguing that organisms do not naturally understand their environment ‘in terms of a geometrical space, independent arrow of time, and Newtonian mechanics’ (Klyubin, Polani & Nehaniv, 2005b, p. 128), but in terms of what it affords them to perceive and do. From this agent-centric perspective, the concept of the environment is a by-product of its embodiment as ‘the interplay between the agent’s sensors and actuators’ (Klyubin, Polani and Nehaniv, 2005, p. 745; and Appx. D).

These theories highlight the importance of maximising potential and perceivable control in human decision-making, but they do not account for the ubiquity of behaviours in the animal kingdom that are deducible from this principle. Polani (2009) offers a two-part explanation based on information theory, theoretical biology and physics. As a first step, he argues that information represents a universal ‘currency of life’. For an organism to survive, it must continually process information, as dictated by information-theoretic bookkeeping laws. According to e.g. one application of Ashby’s (1956) law of requisite variety, an organism can only control their environment in terms of reducing its entropy by a certain amount, if they acquired the same amount of information from the environment beforehand (cf. Touchette and Lloyd, 2000, 2004; and Appx. D). For instance, animals rely on sensing their environment to source the food which maintains their metabolism. Crucially though, information processing also comes with an energetic cost. Sensing for instance requires energy, which must be compensated by the intake of more food. As a consequence, sensing also imposes stronger dependencies of the organism on the environment, which counteracts their chances for survival. The information to be acquired through sensors must thus be traded off with the energy available to an organism (Laughlin, 2001). Polani (2009) argues that the same applies for other information-processing components, including an
organism’s actuators used to cause changes in the environment by inflicting information back. Thus, information can be considered a ‘currency of life’. As second part of his argument, Polani (2009) formulates an information parsimony principle (cf. Polani, Sporns & Lungarella, 2007). It ties to the first part by considering information as ‘currency of life’ under evolutionary pressure: an organism with suboptimal information-processing would waste metabolic energy resulting in lower chances for survival, and would thus be selected against by evolution. The parsimony principle therefore proposes that well-adapted organisms must have efficient information processing in their specific niche. Nature provides plenty of evidence: we find that evolution caused sensor organs, e.g. the eyes of cavefish (Jeffery, 2005), to deteriorate; but it also led to the development of extremely sensitive sense organs operating close to the limits of physics such as the photoreceptors of certain toads, which can even identify individual photons (Baylor, Lamb & Yau, 1979). Crucially, the information parsimony principle translates from the evolutionary to the individual time-scale: an organism equipped with the capabilities for highly efficient sensorimotor processing can only maximise their evolutionary fitness by seeking out niches during their lifetime in which they can potentially leverage these capabilities to the full extent (Polani, 2009). One means to probe the sensorimotor efficiency that a certain situation affords is to measure the maximum amount of information that an agent could pass from its actuator through the environment back into their sensors. This corresponds to an agent’s potential and perceivable control of the environment. Maximising such control yields the behavioural heuristic introduced earlier.

We have referred to a ‘behavioural principle’ and ‘heuristic’ as template from which many specialised behaviours in nature can be derived. These can correspond to intrinsic but also to extrinsic motivations, the latter e.g. in the case of accumulating money or social status. But we can also understand the overarching principle itself as a motivation, which is intrinsic by virtue of its agent-centric formulation of control. In the next section, we formalise an agent’s potential and perceivable control as empowerment, and the corresponding intrinsic motivation as empowerment maximisation.

### 3.2 Informal and Formal Definition

We first define empowerment informally and discuss important properties compared to other quantities. We then establish the necessary formalism to model an agent’s perspective on their world as basis of the empowerment calculation. Based on this, we introduce empowerment in a very general way, and then defend several assumptions to simplify its calculation. We finish with the definition of greedy deterministic EM, one popular variant of formalising the maximisation of empowerment as a model of IM. We define empowerment only for discrete time and state spaces; for continuous versions cf. Jung, Polani and Stone (2011), Mohamed and Rezende (2015), Gregor, Rezende and Wierstra (2017) and Karl et al. (2017), amongst others.

Informally, empowerment measures an agent’s (i) perceivable, (ii) reliable and (iii) potential control over the environment. We say (i) perceivable, because an embodied agent with limited access to their external environment could

---

**Informal Account**

**Empowerment Maximisation**

**Greedy Deterministic EM**

**Discrete time and state spaces**

**Continuous versions**

**Jung, Polani and Stone (2011)**

**Mohamed and Rezende (2015)**

**Gregor, Rezende and Wierstra (2017)**

**Karl et al. (2017)**

---
inflict changes which they cannot distinguish afterwards; empowerment only accounts for an agent’s distinguishable influence on the environment as captured in their future perceptions. The environment is thus treated as by-product of the agent’s embodiment (cf. Gibson, 1979; and Sec. 3.1). In control-theoretic terms, empowerment is thus a combined measure of controllability and observability. Empowerment quantifies (ii) reliability, as it can also take uncertainty in causing different future perceptions into account, which distinguishes it from simpler mobility measures. Finally, empowerment is a (iii) potential quantity, as it measures only the future influence an agent could have without the need to actually exercise this influence.

Crucially, empowerment not only quantifies an agent’s potential control of their immediate future, but the influence they could have on their future perceptions \( n \) steps ahead with action sequences of length \( n \). We call this arbitrary \( n \in \mathbb{N} \setminus \{0\} \) the agent’s lookahead. By default, we always refer to this more generic \( n \)-step formulation.

By maximising empowerment as IR, an agent acts to get into situations which afford the full leverage of their sensorimotor equipment, thus realising the information parsimony principle (Polani, 2009) through behaviour.

For our formalisation, we make an explicit distinction between the objective world that an agent is embedded in, and their beliefs about that world. The perception-action (PA)-loop as introduced in Appx. D represents the first, objective view. The agent’s beliefs about their world are encoded and inferred through a generative model, representing their subjective perspective on the world, mediated by their embodiment. The formalisation of both perspectives draws on joint work in (Biehl et al., 2018). In comparison, we can simplify the formalism slightly thanks to the specific requirements of empowerment.

The PA-loop in our formalisation defines the true state and dynamics of the physical system which includes the agent and the rest of the world. This true state and the true dynamics are usually not directly accessible to the agent. We have originally defined the PA-loop as causal bayesian network (BN) in Appx. D, and we repeat the illustration of the network topology in Fig. 3.1a. To fully specify the loop, we have to define the state spaces of the agent’s sensor \( s \in \mathcal{S} \), memory \( m \in \mathcal{M} \), actuator \( a \in \mathcal{A} \) and the agent-external environment, i.e. the rest of the world \( r \in \mathcal{R} \). We furthermore have to define the (initial) dynamics of the agent’s sensor, memory and external environment, i.e. \( p(s_t | r_t) \), \( p(m_t | s_t, m_{t-1}, a_{t-1}) \) and \( p(r_{t+1} | a_t, r_t) \), respectively, as well as the agent’s policy \( p(a_t | m_t) \). The dot notation marks these dynamics as interventional probability distributions (Appx. A), but for the sake of simplicity, we drop this notation in the rest of this chapter. For a robot, these state spaces and dynamics, except for the policy, would be dictated by physical reality and the robot’s hardware. For a videogame character, they would be given by the game’s mechanics and the character’s setup. For our generic formalisation, we assume them to be arbitrarily fixed, except for the memory state space, the memory dynamics, and the agent’s policy, which we constrain to a specific form.

To allow for more generality, we require a ‘perfect memory’ in which the agent’s complete sensorimotor experience, i.e. all previous perceptions and performed actions, are retained. We thus write the memory state at time \( t \) as \( m_t = (a_{<t}, s_{<t}) \), abbreviating the past sensor and actuator state sequences...
that have affected the memory up to time $t$ with $s_{<t} = (s_0, s_1, \ldots, s_t)$ and $a_{<t} = (a_0, a_1, \ldots, a_{t-1})$, respectively. The memory state space comprises all possible sequences of sensor and actuator values up to time $t$:

$$
\mathcal{M}_t = S \cup \left( \bigcup_{k=1}^{t} S \times (S \times A)^k \right)
$$

(3.1)

The initial memory state is $m_0 = s_0$. We define the memory dynamics via Kronecker’s delta (cf. Eq. A.7) as $p(m_t|s_t, a_{t-1}, m_{t-1}) = \delta_{m_t, (s_t, a_{t-1}, m_{t-1})} \forall m_t \in \mathcal{M}_t, t > 0$, and the initial dynamics as $p(m_0|s_0) = \delta_{m_0, s_0} \forall m_0 \in \mathcal{M}_0$. We drop this requirement of a (perfect) memory later.

To complete the holistic account of the objective agent-environment interaction described by the PA-loop, we only have the agent’s action policy $p(a_t|m_t)$ left to define. From an external perspective, it specifies the probability that the agent picks a certain action, given their sensorimotor experience in memory. To select an action, the agent has to assign reward to individual actions in an action-value function, a process facilitated by their motivational model. For this model to be intrinsic, the underlying IR must, amongst others (cf. Sec. 2.2.3), be computed from an agent’s perspective, based on agent-internal components only. Empowerment as specific IR requires an agent to predict how much control a certain situation affords in terms of the potential impact the agent...
could make with their actions on their future perceptions starting from that situation. This impact is mediated by the agent-external environment.

Crucially though, neither the actual state nor the dynamics of the external environment are directly accessible to the agent – they must be considered latent variables. To calculate empowerment, an agent must thus (i) use their sensorimotor experience to infer the latent state of the external environment as well as the environment and sensory dynamics and, based on that, (ii) predict the consequences of their actions on their future perceptions. This inference and prediction is facilitated via an agent-internal generative model which is encapsulated in the action policy and plugged into the action-value function. It is called generative, because it relates parameters $\Theta$ and latent variables $R$ as ‘generative causes’ to sensor values $S$ as data in a joint distribution.

Similar to the PA-loop, we define the generative model as causal BN $G = (V_G, P_G)$, with the graph structure shown in Fig. 3.1b. We distinguish variables in the generative model from the corresponding, modelled variables in the PA-loop with additional notation: hatted variables, e.g. $\hat{S}$, are assumed by the agent internally, and serve as models for variables in the PA-loop which have not been resolved yet or cannot be accessed directly. The generative model comprises the following random variables unrolled in time:

- Agent sensor $\hat{S}$ with state space $\mathcal{S}$
- Agent actuator $\hat{A}$ with state space $\mathcal{A}$
- The rest of the system $\hat{R}$ with state space $\mathcal{R}$

It furthermore contains continuous parameters $\Theta = (\Theta^1, \Theta^2, \Theta^3)$ and hyperparameters $\Xi = (\Xi^1, \Xi^2, \Xi^3)$, the latter of which we assume to be fixed to $\xi^* = (\xi^{1*}, \xi^{2*}, \xi^{3*})$. To fully specify the model, we first have to define the state spaces of these variables. For successful inference of the parameters, we assume the actuator and sensor state spaces to match, i.e. $\hat{A} = A$ and $\hat{S} = S$. We also assume that $\hat{R} = R$, i.e. the agent knows in principle the possible states of their external environment. The state spaces of the (hyper-) parameters $\Theta$ and $\Xi$ are determined by the choice of $\hat{S}$, $\hat{A}$, and $\hat{R}$.

To complete the definition of the generative model, we have to define the following probability distributions in $P_G$. We distinguish them as models of the true dynamics in the objective PA-loop by writing $q$ instead of $p$:

- Sensor dynamics model $q(\hat{s}_t | \hat{r}_t; \theta^1)$
- Environment dynamics model $q(\hat{r}_{t+1} | \hat{a}_t, \hat{r}_t; \theta^2)$
- Initial environment state model $q(\hat{r}_0; \theta^3)$
- Belief priors $q(\theta^1; \xi^1), q(\theta^2; \xi^2), q(\theta^3; \xi^3)$
- Belief hyperpriors $q(\xi^1), q(\xi^2), q(\xi^3)$
- Action probabilities $q(\hat{a}_t)$

The individual parameters $\theta = (\theta^1, \theta^2, \theta^3)$ encode the agent’s beliefs in the sensor dynamics, in the environment dynamics and in the initial state of the environment, respectively. Because the hyperparameters are fixed to $\xi^*$,
we can define the hyperpriors as Dirac distributions \( q(\xi^i) = \delta(\xi^i - \xi^{i, *}) \) for \( i \in \{1, 2, 3\} \) (cf. Eq. A.8). The agent’s memory is not explicitly represented in the generative model as it is resolved in all previous perceptions and performed actions up to the current time step. Furthermore, the actuator in the generative model has no parents, implying that the agent chooses their actions freely to initiate actual behaviour which manifests in the PA-loop or to probe the consequences of possible future actions. Note that these belief distributions are not interventional (cf. Appx. B). This is because the generative model in Fig. 3.1 already respects the interventional character of actions by considering them root nodes. Also, successful inference requires us not to intervene in sensor states but to use actual observations.

Empowerment is calculated on latent environment states and relies on an agent’s ability to predict the impact of action sequences of length \( n \) on the future perception \( \hat{s}_{t+n} \). We abbreviate such \( n \)-step sequences of random variables as \( \tilde{a}^n_t = (a_t, a_{t+1}, \ldots, a_{t+n-1}) \). The agent can infer the probability of the current and past environment state as well as the probabilities of future action consequences through their generative model. They are given as part of the complete posterior, which more generally predicts the consequences of action sequences on (yet) unobserved variables. For a convenient formalisation of the complete posterior, we identify the following conditional independence assumptions in the generative model by applying the \( d\)-separation criterion (cf. Pearl, 1988; and Appx. B) on the network topology in Fig. 3.1b:

\[
(\hat{s}^n_{t+1}, \hat{r}^n_{t+1} \perp \hat{s}_{t-1}, \hat{A}_{t-1}) \mid \hat{A}^n_t, \hat{R}_t, \Theta
\]  

(3.2)

Based on this, we can write the complete posterior as product of a predictive- and a posterior factor, parametrised by \( \theta = (\theta^1, \theta^2, \theta^3) \) and fixed \( \xi = (\xi^1, \xi^2, \xi^3) \):

\[
q(\hat{s}^n_{t+1}, \hat{r}^n_{t+1}, \theta|\tilde{a}^n_t, m_t; \xi) = q(\hat{s}^n_{t+1}, \hat{r}^n_{t+1}|\tilde{a}^n_t, \hat{R}_t; \theta) \cdot q(\hat{r}_{t-1}, \theta|m_t; \xi) 
\]  

(3.3)

Note that \( \theta^3 \) is only used in the posterior to parametrise the initial environment state. For brevity, we write the agent’s sensorimotor experience in memory as \( m_t = (s_{t-1}, a_{t-1}) \). The complete posterior provides us with estimates of the past but also of future latent environment states \( \hat{R}^{t+1, n}_0 = (\hat{R}_{t-1}, \hat{R}^n_{t+1}) \), and with estimates of parameters \( \Theta \) and future sensor states \( \hat{s}^n_{t+1} \). For our formalisation of empowerment, we do not use the complete posterior explicitly but its two factors. Since open-loop empowerment does not depend on intermediate future sensor states or past latent environment states \( t < t \), we can simplify both factors accordingly. Given the Markov assumption in our generative model, future sensor states only depend on parameters \( \theta = (\theta^1, \theta^2) \) and the environment state \( r_t \), not on previous \( r_{<t} \). We simplify the posterior factor by marginalising out earlier latent environment states:

\[
q(\hat{r}_t, \theta|m_t; \xi) = \sum_{\hat{r}_{<t}} q(\hat{r}_{<t}, \theta|m_t; \xi) 
\]  

(3.4)
We then simplify the predictive factor to retain only the future sensor state \( n \) steps ahead, rather than a sequence of sensor and world states:

\[
q(\hat{s}_{t+n}|\hat{a}_t^n, \hat{r}_t; \theta) = \sum_{\hat{s}_{t+1}^{n-1}, \hat{r}_{t+1}^{n-1}} q(\hat{s}_{t+1}^{n-1}, \hat{r}_{t+1}^{n-1}|\hat{a}_t^n, \hat{r}_t; \theta) \quad (3.5)
\]

\[
= \sum_{\hat{s}_{t+1}^{n-1}} \left[ \prod_{k=1}^{t+n-1} \sum_{\hat{r}_{k+1}^{n-1}} q(\hat{s}_{k+1}|\hat{r}_{k+1}^{n-1}; \theta^1) q(\hat{r}_{k+1}|\hat{a}_k, \hat{r}_k; \theta^2) \right] \quad (3.6)
\]

\[
= \sum_{\hat{r}_{t+n}} q(\hat{s}_{t+n}|\hat{r}_{t+n}; \theta^1) q(\hat{r}_{t+n}|\hat{a}_t^n, \hat{r}_t; \theta^2) \quad (3.7)
\]

In Eq. 3.6 we write the marginalised predictive factor as autoregressive distribution of the sensor and environment dynamics. Since we are only interested in the final sensor state \( \hat{S}_{t+n} \), we simplify this further in Eq. 3.7, using a recursive definition of the \( k \)-step environment dynamics for \( k > 1 \):

\[
q(\hat{r}_{t+k}|\hat{a}_t^k, \hat{r}_t; \theta^2) = \sum_{\hat{r}_{t+k-1}} q(\hat{r}_{t+k}|\hat{a}_{t+k-1}, \hat{r}_{t+k-1}; \theta^2) q(\hat{r}_{t+k-1}|\hat{a}_t^{k-1}, \hat{r}_t; \theta^2)
\]

(3.8)

Each value of e.g. \( \Theta^2 \) as parameter of the dynamics \( q(\hat{r}_{t+k}|\hat{a}_{t+k-1}, \hat{r}_{t+k-1}; \theta^2) \) represents one hypothesis about how the agent’s environment works. Crucially, the predictive factor (Eq. 3.5) needs no updating based on past experience, as it only depends on changes to \( R_t \) and \( \theta \). The posterior factor (Eq. 3.4) in contrast needs updating at different time steps to yield good estimates of the latent environment states and parameters. At time \( t \), this is done by plugging the sensorimotor experience \( m_t = (a_{<t}, s_{<t}) \) as data into the generative model, i.e. \( \hat{A}_{<t} = a_{<t}, \hat{S}_{<t} = s_{<t} \) and computing the posterior. This procedure constitutes what is commonly referred to as Bayesian inference. We have introduced the concept of a posterior factor here to allow for a fair assessment of the simplifying assumptions we make in later experiments, and to provide links for future work. We however do not detail the actual procedure of computing the posterior factor here, but point the reader to our treatment of its exact and variational Bayesian inference in (Biehl et al., 2018).

We now have everything in place to define empowerment formally. By arguing for information as ‘currency of life’, Polani (2009) also advocates information theory (cf. Appx. C) as universal means to formalise efficiency principles across different agent morphologies. Empowerment is such an information-theoretic efficiency principle, which abstracts the substrate of agent-specific embodiment. It corresponds to the maximum causal information flow (Ay & Polani, 2008) from an agent’s actuators \( A^t \) through the latent environment to their future sensor state \( \hat{S}_{t+n} \). Central to its formalisation is the interpretation of the environment as information-theoretic communication.
channel (cf. Touchette & Lloyd, 2000) with actions as inputs and perceptions as outputs. Empowerment $E$ is then given by the channel capacity (cf. Eq. C.24):

$$E(\hat{r}_i; \theta) = \max_{q(\hat{a}_t^i)} I(\hat{A}_t^i \rightarrow \hat{S}_{t+n}|\hat{r}_i; \theta)$$

(3.9)

$$= \max_{q(\hat{a}_t^i)} \sum_{\hat{a}_t^i, \hat{S}_{t+n}} q(\hat{a}_t^i)q(\hat{S}_{t+n}|\hat{a}_t^i, \hat{r}_i; \theta) \log \frac{q(\hat{S}_{t+n}|\hat{a}_t^i, \hat{r}_i; \theta)}{\sum_{\hat{a}_t^i} q(\hat{S}_{t+n}|\hat{a}_t^i, \hat{r}_i; \theta)q(\hat{a}_t^i)}$$

(3.10)

This formalises state-dependent empowerment$^1$, where the channel distribution is conditioned on a specific latent environment state $\hat{r}_i$ and parameters $\theta$, given by the posterior factor (Eq. 3.4). To calculate empowerment, an agent must freely choose the distribution $q(\hat{a}_t^i)$ which maximises the amount of information they could inject into the environment via possible $n$-step action sequences and perceive again with their sensor later. This optimisation problem can be solved by exhaustive enumeration with arbitrary precision using the Blahut-Arimoto algorithm (Arimoto, 1972; Blahut, 1972), or via approximations as in recent joint work (Salge et al., 2018).

While an agent’s empowerment is inherently a state-dependent quantity, we need to assess the expected empowerment of actions as basis of motivation. We define the empowerment action-value function$^2$, as the expectation of the state-dependent $n$-step empowerment $E(\hat{r}_{t+1}; \theta)$ from Eq. 3.9 over all latent environment states $\hat{r}_{t+1} \in \hat{R}$ that a certain action $\hat{a}_i$ could yield. This distribution of states is conditioned on the possible preceding environment states $R_t$ and parameters $\Theta$, obtained from the posterior factor (Eq. 3.4):

$$E(\hat{a}_t, m_t; \xi) = \mathbb{E}_{\hat{r}_{t+1}, \Theta|m_t, \xi}[E]$$

(3.11)

$$= \int \sum_{\hat{r}_{t+1}} q(\hat{r}_{t+1}, \theta|\hat{a}_t, m_t, \xi)E(\hat{r}_{t+1}; \theta)d\theta$$

(3.12)

$$= \int \sum_{\hat{r}_{t+1}} \left[ \sum_{\hat{r}_t} q(\hat{r}_{t+1}|\hat{r}_t, \hat{a}_t, \theta^2)q(\hat{r}_t, \theta|m_t, \xi) \right] E(\hat{r}_{t+1}; \theta)d\theta$$

(3.13)

The calculation of this action-value function is a multi-stage process, and we illustrate it in Fig. 3.2 for 3-step empowerment using coloured, dashed lines. The figure shows the generative model from Fig. 3.1b further unrolled in time, with hyperparameters now set to fixed values $\xi$. In order to calculate the expected empowerment for an assumed action $\hat{a}_t$, an agent must first (i) plug all sensorimotor experience up to time $t$, i.e. $m_t = (s_0, a_0, s_1, \ldots, a_{t-1}, s_t)$, into the model. These sensor and action values have been directly observed and performed, respectively, and are therefore not hatted. Bayesian inference on this data yields the posterior factor $q(\hat{r}_t, \theta|m_t, \xi)$ (Eq. 3.4) over parameters and latent environment states at time $t$. For the expectation, the agent then (ii)

---

$^1$ We can only calculate state-dependent empowerment if $\hat{R} = R$, i.e. the latent environment state spaces match; if this is not the case, the agent can compute context-dependent empowerment (cf. Salge, Glackin & Polani, 2014b) which further abstracts the structure of possible states of the external environment. We capture both cases in the notion of situation-dependent empowerment.

$^2$ This expectation differs from our definition of the action-value function in (Biehl et al., 2018), where $\hat{a}_i$ is incorporated as fixed first action in otherwise freely chosen $n + 1$-step action sequences used to condition the posterior over the future sensor states.
Figure 3.2: Calculation of the 3-step empowerment action-value in the generative model. Hyperparameters have been fixed to $\xi$, and sensorimotor experience $a_{<t}, s_{<t}$ has been included up to $t$ to infer the posterior factor. Empowerment is calculated for each pair of predicted environment states and sensorimotor dynamics parameters $(r_{t+1}, \theta)$ following the execution of $a_t$ (- -). Its calculation requires to maximise the information that can be injected into future perceptions $\hat{s}_{t+4}$ with action sequences $A_{t+1}^3$ (- -).

uses the environment dynamics $q(\hat{r}_{t+1}|a_t, \hat{r}_t; \theta^2)$ under these inferred states and parameters to predict the follow-up states $\hat{R}_{t+1}$ as consequences of $a_t$ (- -). The agent finally (iii) calculates the 3-step state-dependent empowerment $E(\hat{r}_{t+1}; \theta)$ for each of these possible states and parameters, using the predictive factor (Eq. 3.5) as channel distribution. For each such channel, this involves finding the optimal action distribution $q^*(\hat{a}_{t+1}^3)$ that maximises the causal information flow from actuators $A_{t+1}^3 = (\hat{A}_{t+1}, \hat{A}_{t+2}, \hat{A}_{t+3})$ to the sensor state $s_{t+4}$, starting in the specific latent environment state $\hat{r}_{t+1}$ (- -).

We next introduce two simplifying assumptions to our general formalisation of empowerment. The latter assumes a strict agent-centric perspective under which access to the environment is limited. Similar to other IR functions (cf. Biehl et al., 2018), it thus necessitates inference of the latent environment state and dynamics. Empowerment is distinctive from other IRs in that it quantifies intrinsic control, and it is this characteristic which we would like to focus on in our studies. For our experiments to qualify as proof-of-concepts on the use of EM in computational creativity (CC) and game AI, we eliminate inference as potential source of noise. Without inference, there is no need for memory and we consequently represent the objective interaction of an agent with their environment by means of the simplified, memoryless PA-loop in Fig. D.1b. We make the following assumption:

**Fixed Parameters:** We assume that beliefs over the initial environment state as well as sensor and environment dynamics are given or have been acquired by the agent beforehand, and remain permanently fixed to $\theta^* = (\theta^{1,*}, \theta^{2,*}, \theta^{3,*})$ once the agent starts acting according to EM. The parameters are thus delta-distributed, i.e. $q(\theta_i^i|\xi^i) = \delta_{\phi, \theta_i} \delta(\xi^i - \xi^{i,*})$ for $i \in \{1, 2, 3\}$, and we hide them from the distributions, i.e. $q(s_t | r_t; \theta^{1,*}) = q(a_t | \xi), q(\hat{r}_{t+1} | a_t, \hat{r}_t; \theta^{2,*}) = q(\hat{r}_{t+3} | a_t, \hat{r}_t), q(\hat{r}_0; \theta^{3,*}) = q(\hat{r}_0)$. To simplify the examples in Sec. 3.3 of this chapter, we furthermore assume that the agent can fully observe their environment:
Full observability: We assume that the environment is fully observable and that the agent can fully perceive its state, i.e. \( s_t = r_t \). Objectively, this is still modelled by the memoryless PA-loop in Fig. D.1b that accounts for both latent environment and sensor states, but the sensor dynamics become the deterministic mapping \( p(s_t|r_t) = \delta_{s_t,r_t} \). From the agent’s perspective, modelling the latent environment states becomes unnecessary and we thus drop it from the agent’s model distributions. We consequently replace their sensor and environment dynamics with a single distribution \( q(\hat{s}_{t+1}|\hat{s}_t, \hat{a}_t) \). Both the distribution over initial environment states and the posterior factor become obsolete, as the agent can observe their environment directly.

Full observability subsumes the prior assumption that \( \hat{\mathcal{R}} = \mathcal{R} \). We use both assumptions in our applied studies in Sec. 6.5 as well as 7.5, whereas only the latter draws on the second assumption. We discuss limitations caused by these assumptions in Sec. 6.6 as well as 7.6, and the opportunities in relaxing them for future work in Sec. 8.2 and 8.3.

Based on these assumptions, state-dependent empowerment simplifies to:

\[
\mathcal{E}(\hat{s}_t) = \max_{q(\hat{a}_t^n)} I(\hat{A}_t^n \rightarrow \hat{S}_{t+n}|\hat{s}_t) \tag{3.14}
\]

\[
= \max_{q(\hat{a}_t^n)} \sum_{\hat{a}_t^n, \hat{s}_{t+n}} q(\hat{a}_t^n)q(\hat{s}_{t+n}|\hat{a}_t^n, \hat{s}_t) \log \frac{q(\hat{s}_{t+n}|\hat{a}_t^n, \hat{s}_t)}{\sum_{\hat{a}_t^n} q(\hat{s}_{t+n}|\hat{a}_t^n, \hat{s}_t)q(\hat{a}_t^n)} \tag{3.15}
\]

Similar to Eq. 3.8, we write the \( n \)-step sensorimotor dynamics recursively as:

\[
q(\hat{s}_{t+n}|\hat{a}_t^n, \hat{s}_t) = \sum_{\hat{s}_{t+n-1}} q(\hat{s}_{t+n}|\hat{a}_{t+n-1}, \hat{s}_{t+n-1})q(\hat{s}_{t+n-1}|\hat{a}_t^{n-1}, \hat{s}_t) \tag{3.16}
\]

The empowerment action-value function also simplifies to:

\[
\mathcal{E}(\hat{a}_t, \hat{s}_t) = \mathbb{E}_{\hat{s}_{t+1}|\hat{a}_t, \hat{s}_t}[\mathcal{E}] = \sum_{\hat{s}_{t+1}} q(\hat{s}_{t+1}|\hat{a}_t, \hat{s}_t)\mathcal{E}(\hat{s}_{t+1}) \tag{3.17}
\]

Note that these functions are time-invariant or static (cf. Oudeyer and Kaplan, 2007; and Sec. 2.2.3) – they provide us with the same empowerment value if we visit the same state or state-action pair again at a later time.

We can assume that the preceding inference has been perfect; the agent’s model of the environment and sensor dynamics, or their model of the sensorimotor dynamics in the simplified case, would then match the actual distributions in the PA-loop\(^3\). An agent would calculate the same objective empowerment that could be measured by an omniscient observer\(^4\). More generally though, we must assume imperfection and hence a mismatch between the agent’s beliefs and the actual distributions in the PA-loop, and the agent thus calculates an epistemic empowerment which reflects their subjective uncertainty (Appx. A) about the world. The \( q \)-notation covers both possibilities.

---

3 As long as we conceptually separate an agent’s beliefs \( q \) from the objective and inaccessible distributions \( p \), empowerment remains an intrinsic quantity, even if the beliefs are accurate.

4 *Ab initio*, empowerment has also been considered on an evolutionary, rather than behavioural, time-scale, and as an objective, rather than subjective, measure of control. Our assumptions here can thus already be found in early work e.g. by Klyubin, Polani and Nehaniv (2005b).
To define empowerment maximisation (EM) as a model of IM, we need to formalise how empowerment as IR informs action selection in an agent’s action policy. Given our assumptions of full observability, the agent performs actions based on their current sensor reading, which unambiguously reflects the current environment state. The objective policy is thus \( p(a_t | s_t) \), as specified by the memoryless PA-loop. As a subtle point, we distinguish this from the agent’s subjective policy \( q(\hat{a}_t | \hat{s}_t) \), which is based on the present sensor state and an assumed action. As in the majority of related work, we formalise EM as greedy action selection with added stochasticity:

\[
q(\hat{a}_t | s_t) = \begin{cases} \frac{1}{|A^*(s_t)|}, & \text{if } \hat{a}_t \in A^*(s_t), \\ 0, & \text{otherwise.} \end{cases} \quad \text{with } A^*(s_t) = \arg\max_{\hat{a}_t} \mathcal{E}(\hat{a}_t, s_t) \quad (3.18)
\]

The set \( A^*(s_t) \subseteq A \) contains all equally optimal actions \( \hat{a}^*_t \), i.e. all actions that would obtain the same same maximum empowerment action-value if performed in the sensor state \( s_t \). If there is only one such action, it is chosen with certainty, i.e. action selection is deterministic. If there are several, each is selected with the same probability, i.e. they are equally likely to be performed. For a tie, we thus have stochastic action selection on \( A^*(s_t) \). Crucially, EM as a model of IM thus requires a double maximisation: in (i) determining the empowerment reward as maximum causal information flow in Eq. 3.14 and in (ii) choosing the actions which maximise this reward in Eq. 3.18. This distinguishes empowerment from many other models of IM that get by without the first (i) maximisation, and complicates its approximation.

In this section, we have first introduced empowerment as IR informally. We then formalised it based on an agent’s generative model of their environment which requires inference of the environment’s states and dynamics. We simplified the general formalism for this thesis by dropping inference and assuming full observability. We finally defined EM as an action policy based on the empowerment action-value function. These definitions are critical for the rest of the thesis, but they do not necessarily appeal to intuition. In the next section, we foster the reader’s understanding by discussing the properties of empowerment and EM based on examples and formal arguments.

3.3 Properties and Examples

Empowerment is measured in bits of Shannon (1948) information. It is non-negative and increases the more potential, perceivable and reliable control an agent has over their environment. We can understand the nature of this control better by considering a trade-off inherent to the empowerment calculation. It shows when writing information flow (Eq. C.21) as difference of entropies (cf. Eqs. C.15–C.18):

\[
\mathcal{E}(\hat{s}_t) = \max_{q(\hat{a}^n)} I(\hat{A}_t^n \rightarrow \hat{S}_{t+n}|\hat{s}_t) \quad (3.19)
\]

\[
= \max_{q(\hat{a}^n)} H(\hat{S}_{t+n}|\hat{s}_t) - H(\hat{S}_{t+n}|\hat{A}_t^n, \hat{s}_t) \quad (3.20)
\]
For the maximisation in Eq. 3.20, an agent must find the optimal distribution over action sequences which maximises the left-hand and minimises the right-hand term. The left-hand entropy is maximum if each possible future perception is equally likely. The right-hand conditional entropy is minimum if a specific action sequence leads to exactly one sensor state. An agent’s empowerment is thus maximum if their actions can cause a rich set of sensory futures that are yet cleanly distinguishable based on the respective action.

Information flow as above is given by the mutual information on causal probability distributions. If we drop the requirement for causality, we can exploit the symmetry of the mutual information for an alternative interpretation of ‘non-causal’ (NC) empowerment (cf. Mohamed & Rezende, 2015):

$$E_{\text{NC}}(\hat{s}_t) = \max_{q(a_t^n)} I(\hat{A}_t^n; \hat{S}_{t+n}|\hat{s}_t)$$

$$= \max_{q(a_t^n)} H(\hat{A}_t^n|\hat{s}_t) - H(\hat{A}_t^n|\hat{S}_{t+n}, \hat{s}_t)$$

An agent then maximises mutual information rather than information flow. The left-hand entropy in Eq. 3.22 is maximum if all possible action sequences are equally likely to be performed. The right-hand conditional entropy is minimum, if each individual action sequence $a_t^n$ could be inferred retrospectively, given the state $\hat{s}_t$ in which it was originally executed and a specific sensory outcome $s_{t+n}$. Empowerment is thus maximum if an agent can perform as many different action sequences as possible, while keeping them distinguishable given the resulting perception.

While the causal formulation in Eq. 3.20 emphasises the breadth of final sensory states, the non-causal version in Eq. 3.22 highlights the importance of maintaining as many degrees of freedom in action as possible. These observations also hold for the more general, partially observable case: empowerment is zero when an agent believes to have no influence on their environment as perceived through their sensors. This is reminiscent of Seligman’s (1975)’s concept of helplessness: ‘A person or animal is helpless with respect to some outcome when the outcome occurs independently of all his voluntary responses’ (ibid., p. 17). We usually deal with epistemic empowerment, rendering helplessness relative to an agent’s subjective beliefs.

We illustrate empowerment as $\text{IR}$ based on two examples in a gridworld simulation which is discrete in time, has discrete state and action spaces, and affords full observability. In both cases, the environment state is defined in terms of the agent’s absolute position in Cartesian coordinates, i.e. $\mathcal{R} = \mathcal{X} \times \mathcal{Y}$ with $\mathcal{X} = \{x \in \mathbb{N}|0 \leq x \leq r_w\}$, $\mathcal{Y} = \{y \in \mathbb{N}|0 \leq y \leq r_h\}$ and $r_w, r_h = 9$ corresponding to the width and height of the world. The environment state is thus a 2-component vector $r_t = (x, y)$. The agent can furthermore sense their position perfectly, i.e. $\mathcal{R} = S$ and $p(s_t|r_t) = \delta_{s_t, r_t}, \forall s_t, r_t \in \mathcal{R}$. At each time step, they can move in straight directions or idle, i.e. $\mathcal{A} = \{\text{north, east, south, west, idle}\}$. For the first example, we define the (objective) environment dynamics as:

$$p(r_{t+1}|a_t, r_t) = \delta_{r_{t+1}, p(a_t, r_t)}$$

Gridworld
with deterministic environment state transitions given by:

\[
\rho(a_t, r_t) = \begin{cases} 
  r_t & \text{if } a_t = \text{idle}, \\
  (r_{t,x}, \min(r_{t,x}, r_{t,y} + 1)) & \text{if } a_t = \text{north}, \\
  (\min(r_{t,y}, r_{t,x} + 1), r_{t,y}) & \text{if } a_t = \text{east}, \\
  (r_{t,x}, \max(0, r_{t,y} - 1)) & \text{if } a_t = \text{south}, \\
  (\max(0, r_{t,x} - 1), r_{t,y}) & \text{if } a_t = \text{west}.
\end{cases}
\]  

(3.24)

Here, \(r_{t,x}\) corresponds to the \(x\)- and \(r_{t,y}\) to the \(y\)-component of state \(r_t\). The dynamics express that the agent cannot move cross the boundaries of the grid, which we can picture as walls that cannot be penetrated. The dynamics are deterministic, in that for a given environment state, each action leads to exactly one follow-up environment state with certainty.

We assume that the agent’s beliefs match the actual dynamics, and that the agent can fully perceive the current environment state. Since the sensor dynamics are also deterministic, so is the agent’s overall \(n\)-step sensorimotor mapping \(q(\hat{s}_{t+n} | \hat{a}_t^n, \hat{s}_t)\) (cf. Eq. 3.16). Fig. 3.3 shows an agent’s \(n\)-step empowerment at different positions within this gridworld, for lookaheads \(n = 1, 2, 3, 6, 8, 10\). For \(n = 1\), empowerment is very distinct and low at the edges and in the corners. This is because the walls constrain the agent’s mobility: in the top-left corner for instance, the agent can neither move north or west; performing these actions yields the same sensory consequence as idling, and empowerment is thus reduced. For larger lookaheads, we get
higher empowerment values and a smooth gradient. Empowerment is highest in the middle of the arena where the agent can access the maximum number of distinct states with \( n \)-step action sequences.

Our description already indicates that in deterministic scenarios, empowerment becomes a simple reachability measure. More precisely, given deterministic sensorimotor dynamics, empowerment reduces to the logarithm of the number of distinguishable sensor states that can be reached from a specific origin state with the available \( n \)-step action sequences:

\[
\mathcal{E}(s_t) = \max_{q(s_{t+1} | \hat{a}^n_t, \hat{r}_t)} I(\hat{A}^n_t \rightarrow \hat{S}_{t+1} | s_t) = \max_{q(s_{t+1} | \hat{a}^n_t)} H(\hat{S}_{t+1} | s_t) \tag{3.25}
\]

\[
= - \sum_{s \in \hat{S}^n_{t+1}} \frac{1}{|\hat{S}^n_{t+1}|} \log \frac{1}{|\hat{S}^n_{t+1}|} = \log |\hat{S}^n_{t+1}| \tag{3.26}
\]

We explain this result step-by-step. For a deterministic sensorimotor dynamics model \( q(s_{t+1} | \hat{a}^n_t, \hat{r}_t) \), the \( n \)-step action sequences \( \hat{A}^n_t \) resolve uncertainty in the outcome \( S_{t+1} \) entirely. We thus get \( H(\hat{S}_{t+1} | \hat{A}^n_t, \hat{r}_t) = 0 \) and Eq. 3.20 simplifies to the right-hand side of Eq. 3.25. The remaining entropy is maximum for a uniform distribution over sensory futures (cf. Fig. C.2). We can always specify an action distribution \( q(\hat{a}^n_t) \) that yields such a uniform distribution over the set of reachable states \( \hat{S}^n_{t+1} = \{ s_{t+1} \in \hat{S} | \exists \hat{a}^n_t \in \hat{A}^n | q(\hat{s}_{t+1} | \hat{a}^n_t, s_t) \geq 0 \} \). In the deterministic case, empowerment is thus given by the right-hand side of Eq. 3.26, with the maximisation simplified to finding the set of reachable states, rather than a specific action sequence.

Consider 1-step empowerment in Fig. 3.3a: assuming the agent is situated in the bottom-left corner, i.e. \( s_t = \hat{r}_t = (0,0) \), they can only produce the future perceptions \( \hat{S}^n_{(0,0)} = \{ (0,0), (0,1), (1,0) \} \), as \( \hat{a}_t \in \{ \text{west}, \text{south}, \text{idle} \} \) yields the same perception \( \hat{s}_{t+1} = (0,0) \). Their empowerment is thus \( \mathcal{E}((0,0)) = \log 3 \approx 1.57 \). From the position \( (1,1) \) though, they can reach sensory futures \( \hat{S}^n_{(1,1)} = \{ (1,1), (0,1), (1,0), (1,2), (2,1) \} \) and their empowerment is thus \( \mathcal{E}((1,1)) = \log 5 \approx 2.32 \). For larger lookaheads and a suitable position, the agent can expand its mobility further and yield higher empowerment overall: e.g. for 3-step empowerment at \( \hat{s}_t = \hat{r}_t = (3,6) \) in Fig. 3.3c, the agent can reach 25 sensor states, i.e. \( \mathcal{E}((3,6)) = \log 25 \approx 4.64 \).

Our second example in Fig. 3.4 resembles a ‘bridge’ over an abyss and illustrates empowerment for stochastic sensorimotor dynamics and absorbing states. The abyss is in the north and south, i.e. \( R^{\text{abyss}} = \{ x, 0 \} \land (x, r_w), 0 \leq x < r_w \} \) with \( r_w = 19, r_h = 6 \), and the whole gridworld can again be considered enclosed by an impenetrable wall. The sensor and action space as well as the sensor dynamics match the previous example, but the environment dynamics are now stochastic: each action can lead to different follow-up states depending on the strength of a ‘wind’ blowing across the bridge from south to west and vice-versa, depending on the agent’s \( x \)-position. The wind direction and speed is modelled by means of a horizontal cosine wave (Fig. 3.4a). We
Figure 3.4: Example of an agent's n-step empowerment in a stochastic gridworld corresponding to a bridge from east to west. The stochastic environment dynamics depend on the speed and direction of a vertically blowing wind, modelled as cosine wave with two repetitions (a). Plots (b)-(d) show the agent's empowerment at different positions, each plot for a different lookahead. The agent senses their absolute position. They can move north, east, south, west and idle, but cannot penetrate the walls.
This empowerment landscape is slightly noisy because we estimated the sensorimotor dynamics

\[ w(r_x) = \cos\left(\frac{r_x}{r_w + 1}\right) \text{ with wavelength } \lambda = 2 \times 2\pi \]  

(3.27)

The stochastic environment dynamics extend the deterministic dynamics in Eq. 3.24. For \( a_t, r_t \) given, we take the outcome of the dynamics without wind, i.e. \( r^* = \rho(a_t, r_t) \). The stochastic dynamics are then defined as:

\[
p(r_{t+1}|a_t, r_t) = \begin{cases} 
1 & \text{if } r_{t+1} = r_t \land r_t \in \mathcal{R}^{\text{abyss}}, \\
1 - \text{abs}(w(r_{t,x})) & \text{if } r_{t+1} = r^*, \\
\text{abs}(w(r_{t,x})) & \text{if } r_{t+1} = r^{**}, \\
0 & \text{otherwise}.
\end{cases}
\]  

(3.28)

This definition expresses that an agent which is already in the abyss, i.e. \( r_t \in \mathcal{R}^{\text{abyss}} \), remains in this absorbing environment state no matter which action they perform. For any other state \( r_t \notin \mathcal{R}^{\text{abyss}} \), they either transition to the original follow-up state \( r^* \) given by the deterministic dynamics in Eq. 3.24, or to a state \( r^{**} \) shifted by one cell into the wind direction:

\[
r^{**}(r^*) = (r^*_x, \min(r^*_y, \max(0, r^*_y + \text{sgn}(w(r^*_x))))))
\]  

(3.29)

Here, \( \text{sgn}(w) \) corresponds to the sign function, and the resulting state is forced into the interval \([0, r_h]\). The transition probabilities are based on absolute values of the wind direction in \([0, 1]\): the higher the wind speed at a certain coordinate, the more likely the agent is to end up in the shifted state.

Again, we assume full observability and that the agent’s beliefs match the actual dynamics. Figs. 3.4b–3.4d show an agent’s empowerment\(^5\) at different positions on this ‘bridge’ for lookaheads \( n = 1, 2, 4 \). Similar to the first scenario, empowerment close to the eastern and western edges is lower because the walls restrict the agent’s movement. In contrast to the previous example though, we have zero empowerment in the abyss in the north and south. Here, the agent is in ‘free fall’ – they cannot influence the next environment state, and consequently have no control over their future perceptions either. The empowerment for positions on the bridge reflects the stochasticity in the agent’s sensorimotor dynamics. It allows us to differentiate positions in terms of the agent’s risk of being blown into the abyss.

For example, consider the agent’s empowerment at position \( \hat{s}_t = \hat{r}_t = (0, 5) \) and the corresponding wind speed and direction from Fig. 3.4a. Here, empowerment is minimum as \( w(\hat{r}_{t,x}) = \cos(0) = 1 \), i.e. the wind blows full force northbound. For any action, the agent is guaranteed to be blown into the abyss but for \( \hat{a}_t = \text{south} \), which yields the same state \( \hat{s}_{t+1} = \hat{r}_{t+1} = (0, 5) \) with certainty. At position (5,5) in contrast, empowerment is maximum despite the agent standing next to the abyss, for two reasons. Firstly, the wind here blows fully southbound, i.e. \( w(\hat{r}_{t,x}) = \cos\left(\frac{\pi}{2}\right) = -1 \), and the agent’s potential futures are consequently diverse, rather than collapsing in the abyss.

\(^5\) This empowerment landscape is slightly noisy because we estimated the sensorimotor dynamics \( q(s_{t+1}|a_t, \hat{r}_t) \) by sampling from the true dynamics in Eq. 3.28.
moving towards the abyss with $a_t = \text{north}$ yields $\hat{s}_{t+1} = \hat{r}_{t+1} = (5, 5)$, with all other actions resulting in more southern states. Secondly, it is maximum precisely because the wind blows strongest, and action consequences can thus be predicted with certainty. Since the agent’s beliefs about their action consequences match the objective dynamics, maximum empowerment here marks the objectively safest vertical position for the given $x$ coordinate.

Next, we consider the behaviour of agents that greedily maximise state-dependent empowerment as IR. Figs. 3.5a–3.5c show an agent’s trajectory from following the EM policy in Eq. 3.18 for ten time steps in the first gridworld example with deterministic dynamics and for lookaheads $n = 1, 2, 4$. It fades from white to black, with the beginning and end marked by a dot and cross, respectively. In each case, the agent starts in the top-right corner, and moves on the shortest route to the empowerment maximum. The stochasticity of the policy shows in random actions towards equally empowered successor states.

The second example complicates this gradient ascent by its stochastic environment dynamics. Here, an agent’s actions can yield different outcomes with varying probability; perturbed by wind, a move might produce a position with sub-optimal empowerment, given the reachable states. Fig. 3.5d shows two trajectories of length ten, starting from the centre left and right, respectively, for a lookahead of $n = 4$. Here, diagonal lines indicate a horizontal
move where the agent ended up one position further south, compared to performing the same action without wind. Particularly towards the end of the trajectory, the agent has to continuously move against the strong southbound wind and experiences setbacks to maintain the optimum position close to the abyss. They thus leverage their model of the environment dynamics in navigating the gradient, potentially exploiting external perturbations. This example highlights the importance of an accurate model of the sensorimotor dynamics for navigating such an empowerment cliff. In contrast to the first example, a local empowerment maximum not only warrants high reachability, but is also furthest removed from zero-empowerment, absorbing states. In related work, the resulting behaviour is often referred to as ‘death-averse’ (cf. Salge, Glackin & Polani, 2014b). However, in a related publication (Guckelsberger & Salge, 2016) we clarify that at least for living organisms, zero empowerment should rather be associated with helplessness (Seligman, 1975) or considered a proxy for biological death.

As our last example of EM, we reproduce a pendulum balancing task from existing work for its similarity to popular RL benchmarks. Fig. 3.6 shows the same pendulum at increasing time steps. The background illustrates the pendulum’s empowerment at different positions, with the x-axis corresponding to its angle and the y-axis to its angular velocity. The actions are given by positive and negative angular velocities and the solid green line represents the pendulum’s trajectory in this space. The progression from left to right highlights the typical effect of maximising empowerment: the pendulum oscillates towards and eventually balances in the top position from which the largest number of distinct positions can be accessed in a fixed time frame. This simulation is taken from Salge, Glackin and Polani (2013a), who used linearised pendulum dynamics and a discretised action space (Jung, Polani & Stone, 2011). Karl et al. (2017) reproduces the task without these assumptions using a variational, model-based RL empowerment approximation.
Crucially, an agent driven by EM maximises their potential control, without necessarily having to exercise it: the agent’s actual empowerment maximising action policy $q(\hat{a}_t|m_t)$ or $q(\hat{a}_t|s_t)$ can differ from the optimal distribution of action sequences $q^*(\hat{a}_t^n)$ that maximises the channel capacity. In the previous example for instance, an agent does not need to throw themselves into the abyss, although their empowerment incorporates that option. In contrast to e.g. predictive information (cf. Ay et al., 2008; and Sec. 2.2.4), empowerment thus measures the agent’s possible, but not the actual richness of behaviour. This potential information flow differs from the actual information flow caused by the agent’s performance of empowerment maximising actions.

All previous examples and most existing studies are based on open-loop empowerment, where all possible future action sequences are considered in the maximisation of the channel capacity. In closed-loop empowerment in contrast, an agent considers how their policy would affect the probability of performing each action in the sequence. In other words, the closed-loop formulation takes sensory feedback at each future time step along assumed $n$-step trajectories into account. Crucially though, open-loop empowerment can never over, but only underestimate closed-loop empowerment (Capdepuy, 2010). Embedded in EM as motivational model, this can lead to overcautious behaviour: Gregor, Rezende and Wierstra (2017) show that an agent maximising open-loop empowerment does not dare to enter a hazardous area of the world, while closed-loop empowerment maximisation would allow them to navigate better within. Closed-loop empowerment comes with increased computational demands (cf. Salge, Glackin & Polani, 2014b) and most existing work thus focusses on the open-loop formulation. Several approaches have been developed to compute closed-loop empowerment exhaustively and in approximation: Capdepuy (2010) formulates it based on a communication channel with feedback, Salge and Polani (2016) propose to evaluate the empowerment of controllers, and Gregor, Rezende and Wierstra (2017) as well as Binas, Ozair and Bengio (2019) optimise variational bounds on closed-loop empowerment with a model-free RL approach and neural network function approximators. For our models and experiments in Ch. 6 and 7, we exclusively use open-loop empowerment.

According to our working definition (Sec. 2.2.3), EM qualifies as a model of IM. In its general formulation, the empowerment action-value function is agent-centric, in that it is calculated on an agent’s beliefs about the sensory and environment dynamics, based on their generative model of the (objective) PA-loop. Agent-centricity also holds for our simplified version if we assume that these beliefs have been acquired at an earlier time by inference from sensorimotor experience alone. Empowerment is free of semantics because of its information-theoretic formulation. Its calculation only distinguishes the potential sensor states an agent’s actions could yield by their probability and irrespective of their meaning. Furthermore, empowerment is embodiment universal in that it can be applied to any combination of sensors and actuators.

\footnote{The potential nature of empowerment is contingent on the presence of a model of the agent’s sensorimotor dynamics; in model-free approaches (e.g. Mohamed & Rezende, 2015), the agent must necessarily exercise some influence first to estimate their empowerment.}
ber of simultaneous actions on a set of future perceptions. By summarising multiple sensors and actuators into a vector-valued random variable each, they can be treated like single variables in the calculation of channel capacity to produce a scalar reward. Empowerment would be sensitive to the effect of fins on sonar signals, but also to the sensorimotor dynamics induced between propellers and camera images on a drone.

Finally, EM is also likely to realise the diagnostic of open-endedness, operationalised earlier (Sec. 2.2.3) as the ongoing development of skills and knowledge. In order to evaluate open-endedness for the general empowerment formulation, further studies on the interaction of model inference and EM must be conducted. In the meantime, we can only consider behaviour in the absence of model updates. We can show formally that an empowerment maximising agent would neither get attracted to random noise, nor to a dark room. A random noise source by definition cannot be controlled, i.e. different sensory futures cannot be distinguished by actions. The conditional entropy in Eq. 3.20 would reduce to the left-hand entropy, resulting in zero empowerment. An agent in a dark room could not differentiate sensor states in the first place; both entropies in Eq. 3.20 thus become zero, and empowerment vanishes. An agent motivated by EM would only step into a dark room if it served as a passageway to states that warrant more perceivable control. Similar ‘bottlenecks’ feature prominently in our experiments in Ch. 6.

In this section, we have offered more intuitions about empowerment as IR and its maximisation as a model of IM. We have discussed the properties of empowerment and investigated the behaviour of empowerment maximising agents based on formal arguments and example simulations for both deterministic and stochastic sensorimotor dynamics. In the next and final section of this chapter, we briefly sketch the empowerment research landscape and discuss relevant connections to work in physics.

3.4 EMPOWERMENT RESEARCH LANDSCAPE

Since its inception in 2005, empowerment and the maximisation principle have been investigated from many different angles, and related research continues with growing interest. Our goal here is to structure the body of existing studies based on three core hypotheses formulated in early work (cf. Klyubin, Polani & Nehaniv, 2005a, 2005b). We support this by brief, representative examples but defer the detailed discussion of immediately related work to later sections. This provides us with reference points for the contributions that this thesis makes to the overall study of empowerment. We also discuss the connection of EM and causal entropic forces, a similar action principle formulated in physics and motivated from a slightly different angle.

Based on the theoretical motivation and empirical observations summarised in Sec. 3.1, Klyubin, Polani and Nehaniv (2005a, 2005b) have implicitly expressed several empowerment hypotheses, which have been explicitly formulated by Salge, Glackin and Polani (2014b) and inspired many of the following studies. The (i) behavioural hypothesis suggests that adaptation via evolutionary forces led to organisms which, in the absence of specific goals, behave as if they were maximising empowerment. This corresponds to the deduction
of specialised behaviours as exemplified earlier. A stronger version of this hypothesis states that some organisms maximise empowerment directly. It would then correspond to a local, task-independent utility function allowing an agent to increase their fitness in the presence of only sparse evolutionary feedback. Both forms correspond to the application of the information parsimony principle in an organism’s lifetime. Applied to the evolutionary time-scale, the parsimony principle leads to the (ii) evolutionary hypothesis, suggesting that adaptation brought about by natural evolution on average increases the empowerment of evolved organisms. The third (iii) AI hypothesis considers EM a task-independent motivation that can yield goal-oriented behaviours which would otherwise have to be induced via externally imposed rewards. These three hypotheses closely relate to the various incentives to formalise IM (Sec. 2.2.1): the first two are of interest for the realisation of autonomous development in artificial agents and the compliance with and advancement of neuroscience, while the last relates to the goals of increasing task performance and generalisation. Next, we describe representative examples of existing work, and the contribution of this thesis to the three hypotheses.

One means to test the behavioural hypothesis is to demonstrate that EM in simulated agents can yield behaviour that would also be expected in living beings (Salge, Glackin & Polani, 2014b). In his theory of affordances, Gibson famously stated ‘Why has man changed the shapes and substances of his environment? To change what it affords him. He has made more available what benefits him and less pressing what injures him’ (Gibson, 1979, p. 123). Salge, Glackin and Polani (2014a) demonstrate how empowerment maximising agents in a three-dimensional blockworld, capable of digging out and placing blocks, restructure their environment in a way that benefits their specific embodiments: an agent capable of climbing builds a staircase, while a flying agent removes blocks to reduce obstacles. Another agent, cut off by a ‘deadly’ lava stream and unable to fly, builds a bridge over the stream to access a bigger part of the world. Considerable support for the behavioural hypothesis has been provided by experiments where multiple agents interact. As they mutually influence a shared environment with their actions, their individual empowerment can become intertwined. Capdepuy, Polani and Nehaniv (2007) show that two agents that individually maximise empowerment compromise between remaining close and avoiding collision. Similar behaviour has been demonstrated later for continuous, (Salge, Glackin & Polani, 2013b) and for closed-loop, empowerment (Gregor, Rezende & Wierstra, 2017). Capdepuy, Polani and Nehaniv (2007) also show that individual EM in large agent societies yields complex structures such as clusters, membrane-like boundaries and regular patterns as well as pulsing behaviours reminiscent of multi-cellular organisms. In work predating this thesis, Guckelsberger and Polani (2014) have shown how different models to anticipate the actions of other agents affect individual survival strategies in a multi-agent scenario with a scarce energy resource. Although acting as individuals, agents maximise collective survival by consuming greedily, parsimoniously or according to mixed strategies depending on the resource’s scarcity and their anticipation model. This demonstrates the sensitivity of empowerment to an agent’s beliefs about their environment, including other agents, and more generally that EM can trigger biologically plausible behaviours. In Ch. 6, we contribute further evid-
ence to the behavioural hypothesis by demonstrating how EM can yield both supportive and adversarial behaviour in videogame characters that would not be unexpected if exhibited by human players.

A second means to test the behavioural hypothesis is to investigate whether the behaviour of living beings can be anticipated by considering its effects on their empowerment (Salge, Glackin & Polani, 2014b). Trendafilov and Murray-Smith 2013; 2015 have found that control interfaces with reduced empowerment were correlated with feelings of frustration, potentially explaining avoidance behaviour. We discuss this study in more depth in Ch. 7. Our contribution there supports the behavioural hypothesis by establishing links between a person’s empowerment and their reported player experience.

An agent’s empowerment not only depends on the environment state they are in, but also on their capacity to sense and act on that state, i.e. to receive information from, and inject it back into, the environment (Klyubin, Polani & Nehaniv, 2005b). One means to probe the *evolutionary hypothesis* is to adapt an agent’s morphology with the goal to maximise empowerment, and investigate the meaningfulness and diversity of the optimised sensors and actuators (Salge, Glackin & Polani, 2014b). Klyubin, Polani and Nehaniv (2005b, 2008) implement this method by using empowerment as objective function in a genetic algorithm. They evolve both sensors and actuators in a gridworld where a marker substance is emitted from a central source. Possible sensors are characterised by different arrangements of detectors capable of picking up the marker. They find two distinct types of optimal sensors, depending on an agent’s relative position to the source during evolution. For agents close to the centre, the detectors are aligned in a blob and can identify the agent’s absolute displacement from the source. At a sufficiently large offset from the centre though, the detectors re-align into an arc shape that measures the agent’s relative bearing. In contrast to these clearly differentiated sensor layouts, Klyubin, Polani and Nehaniv (2005b, 2008) report a higher diversity of equally optimal actuators. A possible actuator can entail a limited amount of actions to move the agent into different directions. The further away the agent has been placed from the source, the more actions in the optimal actuators have been found to direct towards the source. Both experiments demonstrate that empowerment can be used to evolve meaningful sensors and actuators depending on a specific niche.

Many existing studies support the *AI hypothesis* by showing that EM can yield behaviour which would be considered goal-directed by an external observer, and which would normally have to be induced by extrinsic rewards. For this to hold, the maximisation of empowerment as IR must implicitly align with extrinsic reward achievement (cf. Sec. 2.2.1). Salge, Glackin and Polani (2014a) e.g. observe that an empowerment maximising agent, threatened by a flow of lava in the blockworld environment, shows different ‘death avoidance’ strategies: they e.g. assemble a dam to stop the lava, construct an island to escape it or excavate a tunnel and close it from underneath. Mohamed and Rezende (2015) show similar behaviour for a variational approximation of empowerment via (deep) RL. They furthermore demonstrate in a separate experiment how EM drives an agent to pick up a key and open a door. Again, we get seemingly task-oriented behaviour due to an implicit alignment of these events with an increase in empowerment: picking up the
key provides the agent with the additional option to open the door, and opening the door then warrants additional mobility and thus empowerment. Similar mechanics can be found in many videogames; Anthony, Polani and Nehaniv (2014) show that a bounded-rationality version of empowerment can lead to the identification of sensible strategies in simplified versions of the games Sokoban and Pac-Man, a study which we discuss further in Ch. 5. It has also been shown that EM can solve other benchmarks in RL in an unsupervised fashion, i.e. without relying on extrinsic rewards. One example is the pendulum balancing task mentioned earlier (Fig. 3.6) which has originally been investigated by Jung, Polani and Stone (2011) as well as Salge, Glackin and Polani (2013a). It has more recently been picked up by Karl et al. (2017), who also demonstrated that empowerment can replace extrinsic rewards in the bipedal walker task, where a biped robot with two-joint legs must be balanced, based on torque control and complex sensors.

Extrinsic rewards, e.g. an increase in game score triggered by unlocking a new area, are often very sparse. For IM to replace or complement extrinsic motivation, it must establish a smooth gradient in-between such sparse rewards. In other words, local IRs should ideally correlate with global properties of a system. It has been found that empowerment as local quantity is sensitive to regularities that pervade the whole system. Klyubin, Polani and Nehaniv (2005a) for instance show that empowerment in a maze is anti-correlated with the average shortest distance from one position to any other position, the latter being a global property of the environment. Anthony, Polani and Nehaniv (2008) have studied this relationship between empowerment and global properties of a system further, by considering the environment dynamics as transition graph. They show that in many cases, empowerment as a local quantity can predict the global measure of closeness centrality. We contribute particularly to the AI hypothesis with our contribution in Ch. 6, showing that videogame characters driven by a multi-agent extension of EM realise task-oriented behaviour that would otherwise have to be meticulously hard-coded or learned from manually engineered extrinsic rewards.

Empowerment and the maximisation principle have been motivated by observations and theoretical arguments in psychology and biology (cf. Sec. 3.1). The related concept of causal entropic forces (Wissner-Gross & Freer, 2013) has the potential to root empowerment in physics. It is based on the more fundamental maximum entropy production principle, which has been postulated to arise from first thermodynamic principles (Dewar, 2003, 2005). Causal entropic forces represent a general ‘thermodynamic model of adaptive behavior as a nonequilibrium process in open systems’ (Wissner-Gross & Freer, 2013, p. 1). More specifically, the authors define causal path entropy as a measure of how many paths a system can follow under the laws of physics during a fixed time horizon. At present, causal path entropy is hypothesised to correspond to a proto-empowerment, with its maximisation giving rise to similar self-organisation properties as EM: Wissner-Gross and Freer (ibid.) show in analogy to the discrete gridworld in Figs. 3.5a–3.5c how a particle that maximises causal path entropy is pushed into the middle of a box. Furthermore, the maximisation allows to solve the cart pole problem, a similar scenario as the pendulum balancing task in Fig. 3.6. However, the derivation of the underlying maximum entropy production principle has not been successful
so far (Grinstein & Linsker, 2007), and a connection between empowerment and fundamental physics is therefore yet to be demonstrated.

This concludes our introduction of EM as the central model of IM to be investigated in this thesis. We have motivated empowerment based on observations and theoretical arguments in psychology, biology and physics. We have introduced empowerment and EM informally and formalised it in a general and simplified way for our experiments. Empowerment is calculated on an agent’s beliefs about their world, and we have made the distinction between such models and the objective world particularly explicit. We have discussed the properties of empowerment based on formal arguments and examples, and we have finally outlined related work, structured by three core empowerment hypotheses. In the next two chapters, we build on the preceding overview of different models of IM to systematically review existing work on IR and IM in computational creativity (CC) and game AI.
Part II

SYSTEMATIC REVIEWS
In this chapter, we support the first overarching research question of this thesis, ‘Can IR and models of IM advance CC?’ (RQ.2) by directing the following specific research questions at existing work:

RQ.3 Why have IR and models of IM been used in CC?

RQ.4 How have IR and models of IM been used in CC?

The findings enable us to motivate and contextualise our novel models and applications in Ch. 6 and 7.

Similar to our introduction to IM in Ch. 2, we initially inform these questions through a non-computational angle. In Sec. 4.1, we consider which connections between IM and human creativity have been identified in the field of creativity research. We elaborate on the struggle of defining creativity, and clarify our position. We then answer questions RQ.3 and RQ.4 in Sec. 4.2 through a systematic review of theoretical and applied related work. We constrain the scope of this review by developing a working definition of CC as a cross-disciplinary research endeavour, and applying it together with our working definition of IM models from Sec. 2.2.3. We finally motivate our applied contributions in Ch. 6 and 7 relative to the state-of-the-art.

While there is a surprisingly rich body of research using IR and models of IM in CC, its theoretical grounding is narrow, and it focusses on the benefits of specific models while remaining ignorant of IM as an overarching model class. The main contribution of this chapter is our systematic review of the benefits and applications of IR and IM models in CC, and we distil the findings into two typologies. This represents the first study of its kind, and allows us to advocate the wider use of such model. Our big picture view connects IM research in both CC and creativity studies and thus has the potential to inspire future work within and across both fields.

4.1 INTRINSIC MOTIVATION IN CREATIVITY STUDIES

We inform our review and research questions by drawing on creativity studies, the long-standing, joint effort of psychology, sociology, philosophy, and other non-computational disciplines to investigate the phenomenon of creativity. Creativity studies has been substantially shaped by a tradition in which creativity is considered an exclusively human power (cf. Boden, 1990/2003, pp. 11-14). Most existing research concerns creativity in people, also because of the economic value of creativity for e.g. education and workplace productivity (Guilford, 1950). CC research has been strongly inspired by this anthropocentric1 agenda, adopting the creativity concept as known from creativity studies, i.e. as first and foremost human. The majority of CC research

---

1 We understand anthropocentrism not in terms of human exceptionalism, but as an interpretation of the world in terms of human values and experiences.
focusses on reproducing or augmenting human-level creativity to benefit people, or to learn more about the functional underpinnings of human creativity as a form of computational psychology (Boden, 1990/2003, pp. 283-285). Creativity studies have long recognised the important role of IM in human creativity (cf. Hennessy & Amabile, 2010), and there thus exists a large track record of relevant related work. CC has traditionally translated many important findings from creativity studies into the computational domain, and we want to continue leveraging this heritage for our investigation of IM.

We adopt Plucker, Beghetto and Dow’s (2004) recommendations for good creativity research practice for our investigation into both human and computational creativity. They argue that creativity researchers must:

(a) ‘(...) explicitly define what they mean by creativity,

(b) avoid using scores of creativity measures as the sole definition of creativity (...)

(c) discuss how the definition they are using is similar to or different from other definitions, and

(d) address the question of creativity for whom and in what context’ (Plucker, Beghetto & Dow, 2004, p. 92).

These best practices matter for us twice: they guide our comparison of existing experimental results on the relationship of IM and people’s creativity in Sec. 4.1.2, and they inform the design and evaluation of our own computational studies in Ch. 6 and Ch. 7. They also highlight a critical issue: similar to the overarching study of (artificial) intelligence, the very notions of ‘creativity’ and what it means to be ‘creative’ are notoriously hard to define and operationalise. Before discussing existing studies, we make this challenge transparent and identify a ‘standard definition’ of creativity as a starting point to think about creativity throughout this thesis.

4.1.1 Defining and Operationalising Creativity

Creativity is a very old and highly contested concept; in 1988, Taylor identified 50 different definitions of creativity proposed over the previous five decades, with more definitions added to date. Still and d’Inverno (2016) argue that this ambiguity partly comes from the fact that our contemporary understanding of creativity is based on two different historic traditions. The older tradition dates back to Pagan times, in which the Latin *creare* was identified in the unfolding and dissolution of natural processes, such as the birth of a child or the growth of a plant. This *creare* translates to ‘having an impact through natural forces’ (ibid., p. 149). It is different from facere which is used in early versions of the Latin bible, and corresponds to ‘make out of available materials’ (ibid., p. 149). The second, Christian tradition dates back to St Jerome’s late 4th century work on the Vulgate bible. He replaced the word facere with *creare* to emphasise the ‘creation out of nothing but ideas in God’s

---

2 We venture beyond this inherent anthropocentrism by advocating other perspectives on (computational) creativity, e.g. in Guckelsberger, Salge and Colton (2017). Yet, we consider it important to embrace creativity studies since findings in the study of human creativity can likely be generalised to understand creativity more generally in living beings and machines.
mind’ (Still & d’Inverno, 2016, p. 149). This younger creare is a mixture of the Pagan creare and facere, and means ‘to bring about by making’ (ibid., p. 149). Rather than a property of natural processes, the Christian creare is considered an individual power. In contrast to the universal Pagan version, it has specifically been used to describe the creative genius in art relying on novel ideas, in contrast to craft which was thought to be based on skill alone (ibid.).

We can moreover consider the Greek word δημιουργία (dimiourgos), formed of δῆμος (demos), i.e. ‘the people’ or ‘the public’, and ἔργον (ergon), which means ‘work’ (Babiniotis, 2019). It was used to signify the work of people who engage in public affairs. The ‘creator’, δημιουργός was distinguished from ἄναψας (anafos), the ‘ordinary worker’ (ibid.). This distinction hence emphasises the social impact of creation.

The advent of modern creativity research is marked by Guilford’s 1950 presidential address to the American Psychological Association, in which he stressed creativity as an important but neglected subject of study, with a projected impact on business, education and society as a whole. The address marked a tremendous growth in research effort, resulting in more than 9000 related studies until 1998 alone (Runco, Nemiro & Walberg, 1998).

In a more recent review, Hennessey and Amabile (2010) re-emphasise the study of creativity as a ‘basic necessity’ to ‘make real strides in boosting the creativity of scientists, mathematicians, artists, and all upon whom civilization depends’ (ibid., p. 570). Despite the large body of theoretical and empirical research, the very definition of creativity is still debated. Still and d’Inverno (2016) argue that this is partly the case because Guilford has ‘unwittingly amalgamated’ (Still & d’Inverno, 2016, p. 149) the two [Pagan and Christian] traditions of thinking about creativity in a single notion: the English words creativity and creative now interchangeably refer to both traditions.

Additional complexity arises from the fact that creativity can be considered from at least four different perspectives, identified seemingly independently by Rhodes (1961), Mooney (1963) and potentially others (cf. Jordanous, 2016). We can use the word ‘creative’ to describe a person, such as the designer of a videogame. Furthermore, we can consider the creative process as the individual steps this person engages in to be creative, such as the invention of a new game mechanic based on blending two existing ones. We can then consider the creative product as the outcome of this process, e.g. the finished game. Finally, the press represents the environmental determinants of creativity, such as the game studio. Still and d’Inverno (2016) point out an anomaly in this usage of ‘creative’, as it can be applied to a person, process and product, but also describes the result of the overall system of all four entities. The person and product perspective echoes more strongly the Christian tradition, in which an independent individual as a creative genius operates distinctly from nature, and in which the resulting artefact can eventually be considered separate from its creator. Similarly, process and press relate to systems theories of creativity (e.g. Vygotsky, 1930/1971), which originated from the Pagan tradition, and identify creativity in the ongoing interaction between an individual and their environment. Most researchers nowadays agree that creativity does not happen in a vacuum but is contingent on an individual, historic and societal context. While adding to the conceptual complexity, these
The *four P’s of creativity* represent a useful tool applied throughout (computational) creativity research, and we also adopt them for this thesis.

Systems theories allow us to consider creativity beyond the limitations of an individual, but embedded in a society of agents. This societal perspective is the foundation for Boden’s (2003) epistemic distinction between ‘psychological’ or ‘personal’ *p-creativity*, and ‘historical’ *h-creativity*, respectively: While an individual may produce a process or product that is novel for them personally, the very same process or product may not be novel with respect to the history of creative activity that is co-produced and shared by the society of agents (ibid., p. 1, 43 ff.). Boden calls the first type ‘psychological’, as she deems it central to understanding the psychology of creativity. Wiggins et al. (2015) note that personal or historical novelty (and value) are relational properties between observers, the created artefact, the creator, and the specific context. The distinction between *p-* and *h-creativity* would thus be too simplistic if not appealing to the 4 *P’s of creativity* (e.g. Rhodes, 1961).

The growing wealth of definitions prompted researchers to express that the concept of creativity has ‘almost ceased to mean anything’ (Batey & Furnham, 2006, p. 357). In order to support rather than quench the discussion, Jordanous and Keller (2016) suggest considering creativity an *essentially contested concept* (Gallie, 1955), which ‘inevitably involves endless disputes about their proper uses on the part of their users’ (ibid., p. 169) and for which a fixed ‘proper general use’ (ibid., p. 167) is elusive. While being ‘internally complex’ (ibid., p. 171) in nature, essentially contested concepts are crucially ‘amenable to being broken down into identifiable constituent elements of varying relative importance, dependent on a number of factors such as context and individual preference’ (Jordanous & Keller, 2016, p. 6). It may appear contradictory that other authors simultaneously advocate a ‘standard definition of creativity’ (Runco & Jaeger, 2012). However, this proposal is compatible with the previous position as it rests on the observation that creativity has been defined in terms of the same two underlying factors or components across a large range of studies. Originally put forward by Stein (1953), the definition requires a *product* or *process* to be both *novel* and *useful* in order to be deemed *creative*. Plucker, Beghetto and Dow (2004) have fused this definition with the *four P’s of creativity*:

> ‘Creativity is the interaction among *aptitude*, *process*, and *environment* by which an individual or group produces a perceptible product that is both *novel* and *useful* as defined within a social context’ (Plucker, Beghetto & Dow, 2004, p. 90, emphasis added).

With its two components, the ‘standard definition’ echoes Sternberg’s (1999) claim that ‘the essence of creativity cannot be captured in a single variable’ (ibid., p. 83). But also this two component model can only capture this

---

3 Jordanous and Keller (2016) also argue that the ambiguity in the creativity concept can be captured by means of *family resemblances*, an analogy proposed by Wittgenstein (2009) to understand different uses of the same word. Wittgenstein suggests that there is no core meaning to a word, but that different meanings form a family, ‘a complicated network of similarities overlapping and criss-crossing’ (ibid., §66). We consider this proposal problematic given the so-called ‘standard definition of creativity’: while *value* has sometimes been deemed non-essential, it is currently not clear if creativity can be considered without *novelty*. This would give the word a core meaning and thus make it too specific for a *family resemblance*. Proponents of the *four P’s of creativity* (e.g. Rhodes, 1961) may respond that the ambiguity of creativity only highlights the value of considering it from multiple perspectives.
essence in an abstract sense, since the meaning of ‘novelty’ and ‘value’ varies depending on who evaluates the creativity of what, and in which context (Wiggins et al., 2015; Silvia, 2018). For example, most videogames would be deemed valuable because they are fun to play. However, other games may be valued because they also have a utility beyond the game itself, e.g. ‘serious games’ (Marsh, 2011). Art-house games in contrast might not be fun and yet considered valuable because they are thought-provoking or aesthetically pleasing.

Since creativity is such a multi-faceted concept, Plucker, Beghetto and Dow (2004) advise not to measure it based on scores alone. For our related work analysis, this raises the question of how creativity is commonly operationalised in creativity studies. Kaufman (2016) summarises that ‘there are many different ways in which someone can be creative, and there are almost as many different ways that people try to measure creativity’ (ibid., p. 9). Each of the four P’s of creativity gave rise to different operationalisations, which are further diversified through the application of different measurement instruments. Based on a survey of existing studies, Batey and Furnham (2006) distinguish divergent thinking tests, attitude, interest, personality and biographical inventories, judgements of products, ratings of eminence, self-reported creative activities, as well as ratings by peers, teachers and supervisors.

For this thesis, we do not go so far to consider creativity essentially contested, but we admit that the concept is under-specified, in that its precise definition depends on resolving the evaluation viewpoint and context (e.g. Wiggins et al., 2015; Silvia, 2018), and because additional factors beyond novelty and value may play a role (Jordanous & Keller, 2016). We thus do not treat creativity as a clear-cut concept but use the ‘standard definition’ (Runco & Jaeger, 2012) to capture most people’s understanding to some extent, while admitting other views. We also use this definition as the underpinning for our study design, but follow Plucker, Beghetto and Dow’s (2004) best practices and operationalise creativity qualitatively. Next, we leverage our knowledge on the diverse definitions and operationalisations of creativity to understand and summarise the key findings of creativity studies on the relationship of IM and creativity in people. These insights serve as a preparation for our systematic review in Sec. 4.2.2, in that they highlight why and how computational models of such motivation can advance CC.

4.1.2 Intrinsic Motivation and Creativity

There exists a trivial relationship between motivation and creativity (as in the ‘standard definition’) in that novelty and value in the process or product are contingent on action, and motivation is at the basis of action-taking (Amabile, 2018). However, psychologists noticed early on that the type of reward (cf. Sec. 2.1.1) underlying motivation can have a major effect on creativity: Writers report to be most creative when engaging in writing for its own sake, while externally imposed deadlines or prizes are well known sources of the ‘writer’s block’ phenomenon (Amabile, 1985). Similarly, creative luminaries such as Einstein complained about the detrimental effect of external pressure through academic exams, lessons and instructions on their studies (Amabile, 1979; citing Schilpp, 1949, p. 17). Consequently, understanding the relationship
At this point, the importance of IM for people’s creativity is widely acknowledged, both on theoretical and empirical grounds. Several theories understand IM as a ‘primary motivational precursor to creativity’ (Liu et al., 2016, p. 238); the componential theory of creativity, for instance, embraces IM as a core mechanism underlying individual creativity, next to domain-relevant expertise and creative thinking skills (Amabile, 1983, 2012). This relationship is commonly explained as follows: ‘When people are intrinsically motivated, they will delve into their work and spend more time and effort to collect novel information, understand problems, and generate creative solutions’ (Liu et al., 2016, p. 238). The philosopher Kieran (2014) argues that intrinsically motivated people are more likely to be creative because they can ‘envisage different possibilities, and be directed (...) toward realizing the inherent values of a given domain’ (ibid., p. 7). These theories thus consider the capacities of IM to yield deeper task engagement, supported by exploratory behaviour and the identification of a task’s inherent rewards as facilitators of creativity. They thus follow closely the definition of IM as engaging ‘in an activity for the interest and enjoyment of the activity itself’ (Liu et al., 2016, p. 242; and Sec. 2.1.2), and focus on exploration as the most investigated type of intrinsically motivated behaviour. Martindale (1990) considers exploration in the form of novelty search as the key motivation in artistic and literary creativity, and as a predictor of artistic change, like in the evolution of style.

These and similar theoretical claims, however, clash with inconsistencies in empirical studies. In an effort to reconcile past findings, researchers have conducted a number of meta-analyses. They seek to explain the present ambiguity by identifying and controlling factors that might mediate or moderate the IM–creativity relationship. de Jesus et al. (2013) suspect the use of different conceptualisations of and perspectives on creativity (Sec. 4.1.1) as sources of inconsistency. Their meta-analysis focusses on artistic creativity, and only considers studies that measure creativity in the product using variations of the ‘standard definition’, i.e. with respect to the product’s novelty (or originality), as well as value (or usefulness, or appropriateness). Eligible studies must furthermore employ some measure of IM, and report a Pearson product-moment correlation $r$ on the IM–creativity relationship. These individual correlations are then aggregated into a sample-size weighted mean correlation $\bar{r}$ to account for differences in the number of participants per study, with $N$ corresponding to the total number of participants across all studies. Based on 26 independent experiments in 15 English studies dating from 1990-2010, they find a moderate, positive association of IM and creativity in the product ($N = 6435, \bar{r} = 0.3, 95\% CI = [0.22; 0.37]$). This effect however is not restricted to the product perspective; based on an informal comparison of studies, Malik and Butt (2017) also support a positive relationship between IM and creativity in the process. While de Jesus et al. (2013) have considered arguably few studies to investigate the effect of IM on artistic creativity, Liu et al. (2016) evaluate the effect of IM on workplace creativity in a larger meta-analysis. They also

---

4 Psychologists distinguish different extrinsic motivators such as expected extrinsic reward, external evaluation, externally imposed constraints, competition, etc. We consider these under the same umbrella of positive or negative extrinsic reward.
require studies to measure IM, and to assess creativity based on a variation of the ‘standard definition’. In this workspace setting, the novelty and potential usefulness of employees’ ideas are evaluated with respect to their organisation via self-reports or external assessment. Based on 68 experiments from 63 studies in English, Chinese and Korean dating from 2001-2015, they find that IM makes a moderate contribution to creativity ($N = 19695$, $r = 0.34$, 95% CI = [0.21; 0.46]), thus adding to Jesus et al.’s (2013) findings. Crucially, this relationship holds while Liu et al. (2016) control for other motivational effects that may have an impact on creativity, including extrinsic motivation.

Similar to the psychological account of IM (Sec. 2.1), the previous insights are limited by creativity studies’ methodology, in particular with respect to the functional underpinnings of the IM-creativity relationship. Still, they provide us with essential cues to investigate these underpinnings further in artificial systems. Next, we leverage these cues to analyse how IR and IM models have already, and can further, advance computational creativity.

4.2 REVIEW OF INTRINSIC MOTIVATION IN COMPUTATIONAL CREATIVITY

The notion of computational creativity (CC) describes a research field and a community. It is under constant debate by those who associate their work with it, and at the same time unheard of by many whose work is immediately related. In Sec. 4.2.1, we put forward a more inclusive working definition of CC by reflecting on previous attempts to both delineate it from, and reconcile it with other scientific endeavours. In Sec. 4.2.2, we then work out how IM can benefit CC through a systematic review of related work informed by our working definitions of CC and IM (cf. Sec. 2.2.3). In Sec. 4.2.3, we use these insights to motivate and contextualise our contributions in Ch. 6 and 7.

4.2.1 Computational Creativity: A Working Definition

The study of creativity in a computational context pre-dates modern AI research (cf. Boden, 2015), and has been intimately coupled to it since its inception: ‘Randomness and Creativity’ was one of the topics proposed for the 1956 ‘Dartmouth Summer Research Project on Artificial Intelligence’, now regarded as the founding event of AI (cf. McCarthy et al., 1955/2006). The development of what was first known as artificial creativity5 (Dartnall, 1993; Elton, 1995) has been strongly supported by Boden (e.g. 1992), who put forward a theoretical framework for the study of creativity in AI.

Artificial creativity eventually gave way6 to computational creativity (CC). Adopting an earlier account by Wiggins (2006a), Colton and Wiggins (2012) put forward a working definition of CC as ‘the art, science, philosophy and engineering of computational systems which, by taking on particular responsibilities, exhibit behaviours that unbiased observers would deem to be creative’ (ibid., p. 21). By relying on the judgement of an observer, Colton

---

5 Not to be confused with Saunders and Gero’s (2001) homonymous systems model of creativity.
6 The details of this transition and the exact relationship between the research agendas of artificial creativity and computational creativity are yet to be determined.
and Wiggins bypass the need to define creativity. But this arguably most popular definition has not remained unrivalled: Jordanous (2012) for instance criticises it for emphasising the challenge ‘to engineer a system that appears to be creative to its audience, rather than engineering a system that possesses a level of creativity existing independently of an audience’s perception’ (ibid., p. 250, emphasis added). She later juxtaposes it with an observer-independent definition of CC as ‘the modelling, simulating or replicating of creativity computationally’ (Jordanous, 2016, p. 194). Pérez y Pérez (2015a) emphasises the use of CC as a means to better understand human creativity, by defining it as ‘the study of the creative process employing computers as the core tool for reflection and generation of new knowledge’ (ibid., p. 31).

Some of the variability in these definitions can be explained by CC’s interdisciplinary scope, spanning psychology, cognitive science, mathematics, engineering, computer science (Ackerman et al., 2017), philosophy (McGregor, Wiggins & Purver, 2014), sociology, art, art history, and other disciplines. Besold (2016) even permits CC research to be independent of a computational context and thus blends the boundaries between CC and creativity studies (Sec. 4.1), precisely because this reflects the CC research practice. On the contrary, Veale, Cardoso and Pérez y Pérez (2019) write that CC ‘adopts an explicitly algorithmic perspective on creativity, and seeks to tie down the study of creative behavior to specific processes, algorithms and knowledge structures’ (ibid., p. 2). Ultimately, this diversity in both definitions and research practice reflects the CC research landscape at different times and the varied goals of the community of researchers that contribute to it.

To capture this variability, different instruments to distinguish the various goals within CC have been developed. Boden (1990/2003) marks the two ‘projects’ of ‘understanding human creativity’ and ‘producing machine creativity’ (ibid., p. 1). Similarly, Veale, Cardoso and Pérez y Pérez (2019) identify two perspectives on CC. From the scientific perspective, researchers look for insights into the phenomenon of (human) creativity and the ultimate capabilities of creative people and machines by means of computational modelling and empirical studies. The engineering perspective in contrast focusses on building working systems that embody these theoretical insights, usually to please and benefit people. The ultimate goal here is to engineer a system which can be considered ‘creative in its own right’ (Colton, 2008, p. 6). Ideally, both perspectives are brought together in a ‘symbiotic relationship (...) wherein the artifacts that are produced also serve as empirical tests of the adequacy of scientific theories of creativity’ (Veale, Cardoso & Pérez y Pérez, 2019, p. 1). Similarly, Pérez y Pérez (2018) proposes a cognitive and an engineering perspective as poles of a CC research continuum. These two perspectives closely relate to the diverse goals of CC researchers, summarised by Colton, Charnley and Pease (2011) and Pease and Colton (2011). While the cognitive perspective wants to provide (i) insights into the nature of creativity, efforts from the engineering perspective can be differentiated further based on a system’s level of creative responsibility (Colton & Wiggins, 2012): from (ii) creativity support systems designed to foster human creativity, over (iii) co-creative systems taking

---

7 The study of human creativity through CC has been proposed by Boden (1992), re-emphasised by e.g. Colton and Wiggins (2012), but only put in a definition by Pérez y Pérez (2015a).
on some creative responsibility in interaction with others, to (iv) fully autonomous creative systems. These two CC perspectives are different from the 4 P’s of creativity (cf. Sec. 4.1.1); researchers have additionally advocated the study of CC in the product and process (Colton, 2008), as well as through the agent as a producer and the press as their socio-cultural environment (Jordanous, 2016).

Despite their long-lasting connection, the relationship between AI and CC is highly ambivalent. Since creativity draws on many types of intelligence, Colton and Wiggins (2012) have proposed CC as a final frontier of AI research. But the same and other researchers have also hedged CC from AI more generally: we believe that at present, a large body of work across AI that is closely related to CC’s goals is not captured by CC’s prevailing self-conception. Researchers in AI sub-disciplines like artificial life, robotics or machine learning are presently either completely unaware of CC, or reluctant to embrace the research heritage accumulated under this term (cf. Besold, 2016; Cook & Colton, 2018a). Relevant work is consequently often done in mutual ignorance – arguably even more so from AI more generally – which holds up progress in individual projects and in CC research overall. In our review, we actively counter this situation by incorporating and linking related work beyond the present boundaries. In an effort to disambiguate the CC concept and to formulate a more inclusive working definition for this thesis, we thus relax some prevailing assumptions on (i) the types of creativity considered and modelled, (ii) CC’s relationship to (creative) problem solving, and its inherent (iii) anthropocentrism. We briefly clarify our view on each point.

Types of Creativity

Most research under the heading of CC focusses on what people commonly consider major creative achievements, e.g. artistic acts like poetry, storytelling, musical composition, visual arts and design (Colton & Wiggins, 2012), but also scientific acts such as theory formation in mathematics (Loughran & O’Neill, 2017). Related work is usually domain-specific and focusses on generating a product or artefact that excites people by matching or surpassing the creativity of eminent human individuals. Creativity studies denote such major creative acts as big-c(reativity) (Kaufman & Beghetto, 2009). But there are also little-c, everyday creative acts, from which learning and personal discovery processes are delineated as mini-c. Boden (1990/2003) reminds us that ‘creativity enters virtually every aspect of life’ and is ‘grounded in everyday abilities’, meaning that ‘every one of us is creative, to a degree’ (ibid., p. 1). At present though, the little/mini-c types of creativity only play a minor role within CC, with notable work e.g. in developmental models of early creative behaviour (Aguilar & Pérez y Pérez, 2014, 2015, 2017), and in cognitive architectures implementing theories of human everyday creativity (Sun & Helie, 2015; Wiggins, 2018). But a stronger emphasis on little/mini-c may well enrich CC: the corresponding models usually focus on the individual actions that constitute a creative process, rather than on generating a specific product. They are more domain-general, and can thus serve as components in complex, big-c systems. Moreover, Wiggins et al. (2015) argue from the cognitive perspective that everyday creativity can help us understand the evolutionary basis of creativity. But despite these arguments and exemplary work, little/mini-c creativity receives little attention within CC, and related research is conducted in other AI disciplines. For instance, videogame design as a big-c act with a focus on the final product is considered an important part of the CC research
agenda, but general game-playing as a little-c act with a stronger focus on individual actions in the creative process, is often not\(^8\). Yet, general game-playing is intensely researched as a benchmark in machine learning. In line with e.g. Wiggins et al. (2015) and Besold (2016), we believe that it would be beneficial for CC research to expand its scope beyond big-c acts. When referring to CC in this thesis, we consequently denote the study of all types of creativity, thus embracing work outside the present CC canon and across all areas of AI.

Colton and Wiggins (2012) distinguish CC and AI more generally by stating that AI projects adhere to a problem-solving, but CC follows an artefact generation paradigm, ‘where the automation of an intelligent task is seen as an opportunity to produce something of cultural value.’ (ibid., p. 22). We need to clarify this distinction for two reasons. Firstly, it obfuscates essential commonalities between CC and other areas of AI. Secondly, the artefact generation paradigm can undermine the role of the creative process, and consequently shifts attention away from types of creativity that are less likely to be understood as a final product or artefact: the successful playing of a previously unknown game involves a series of little/mini-c acts, and yet, it is more appealing to consider the resulting game state as product of these actions, than understanding the individual actions that caused this product as artefacts in time and space. If we focus on the notion of artefacts only, the process that gave rise to them is easily forgotten.

We understand that the notion of problem-solving appears at odds especially with artistic creativity; it seems ‘inappropriate’ (ibid., p. 22) to consider the painting of a portrait as a problem to be solved. But this was probably not Colton and Wiggins’ core concern; they rather introduce the notion of artefact generation to make both a methodological and a formal distinction. They firstly consider CC to be more closely related to the methods of artificial general intelligence research, in contrast to the reductionist AI research of the 1980s and 90s studying intelligence as the solution of specific problems using specialised methods (Wiggins, 2018). Secondly, they likely say artefact generation to differentiate a specific form of problem-solving that is typical for artistic creativity. Jennings (2012) labels this as place search, where the end point is not known in advance and we must find the most desirable state given some evaluation criteria. Wiggins’ later observes that in CC, ‘often there is no “solution” and even no [specific] “problem”’ (Wiggins, 2018, p. 3, emphasis added). The path towards this final state or product can be used to assure observers of a system’s creativity (Colton, 2008), but, apart from that, it is often treated as a means to an end. Jennings (2012) further differentiates path search where the final state or product is well known and the goal is to find a novel and valuable (e.g. short) path to reach it. This characterises instances of e.g. scientific creativity such as automated theorem proving or protein synthesis, and is more closely related to traditional ideas of problem-solving as e.g. ‘transforming a given situation into a desired situation or goal’ (Simon, 2001, p. 674). General game-playing serves as an example where both search types blend together: there is an element of place search in identifying game states that afford high score, but it is usually the path to these goal states that is of

\(^8\) Zook, Riedl and Magerko (2011) and Liapis, Yannakakis and Togelius (2014) highlight the importance of creativity for gameplay, but this association has barely been embraced in theoretical or applied CC research (cf. Ventura, 2016a), as also criticised by Moffat (2015).
interest. We can thus understand both artistic and non-artistic creativity as problem-solving: the first as finding a novel and potentially valuable artefact, stressing artefact generation, and the latter as discovering previously unknown and potentially valuable paths to known artefacts, emphasising path discovery.

While we can understand creativity as problem-solving, not all cases of problem-solving require creativity. We clarify this by differentiating creative problem-solving from the more common notion. In his creative systems framework, Wiggins (2006a, 2006b) formalises Boden’s (2003) model of creativity as search on a ‘conceptual space’ of artefacts. Traversing the space with a search strategy can lead to the discovery of new and valuable (partial-) concepts or artefacts. Wiggins (2006b) notes that any form of artefact generation can be solved by an exhaustive search strategy, albeit subject to the halting problem (Turing, 1936). Crucially though, he believes that ‘a human artist who produces valued outputs by exhaustive enumeration is generally less likely to be heralded as an important artist than an artist who is capable of using his or her own heuristics to arrive directly at a valued artefact’ (Wiggins, 2006b, p. 220, emphasis added), and applies the same reasoning to machines. Wiggins thus draws a connection between the search strategies employed and the attribution of creativity. This complements the psychological theory by McGraw (1978), who distinguishes tasks with respect to the potential creativity they require to be solved. He proposes a continuum between an algorithmic solution for which participants know the necessary steps to take beforehand (e.g. standard multiplication), and a heuristic solution, where these steps, i.e. the solution algorithm, have to be discovered by the participant (e.g. functional fixedness problems, for which a solution requires the non-standard use of known objects). Translated to search, heuristic solutions would be less likely to match an a priori, independently known search strategy. Consider the toy examples in Fig. 4.1: The first maze (Fig. 4.1a) can be solved by the ‘wall follower’ strategy which is not necessarily known but, once derived, can be applied to solve any similar maze. However, there exists no similarly simple, yet generalising strategy for solving variations of the second maze (Fig. 4.1b). For people and any rationally bounded agent under time constraints, an exhaustive search strategy is not feasible. If we discard this option, tasks with such a heuristic solution require the development of a non-exhaustive, heuristic search strategy, which makes McGraw’s and Wiggins’ accounts align. We thus distinguish creative problem-solving from the more general concept by associating it with heuristic solutions and search strategies.

We also note that solving problems creatively often requires finding not one, but many solutions: many researchers modelling artistic creativity are interested in discovering a variety of novel and valuable artefacts from the same initial conditions. Fittingly, Guilford (1967) considers creativity as divergent thinking, i.e. the ‘generation of information from given information, where the emphasis is upon variety and quantity of output from the same source’ (ibid., p. 213). We conclude that different types of creativity, both from a product and process perspective, can be understood as specific forms of problem-solving. By making the overlap between CC and problem-solving in AI explicit, we emphasise the differences but also the intimate connection between the fields:

---

9 An unfortunate choice of terminology which yet does not diminish the value of the distinction.
while not all of AI research touches on questions of creativity, a substantial part does, although often without explicit reference.

As our final point, we note that the earlier definitions of CC are all grounded in our human conception of creativity. The cognitive perspective (Pérez y Pérez, 2015a; Jordanous, 2016) focusses explicitly on simulating human creativity, and this focus is also present in the engineering perspective, albeit more implicit: Colton and Wiggins’s (2012) definition for instance implicitly assumes observers to be human, and thus constrains CC’s scope to what is recognised as creative by people. But people’s understanding of creativity is primarily shaped by the observation of creativity exhibited by other people. While this human bias still dominates CC (cf. Veale, Cardoso & Pérez y Pérez, 2019), some researchers have called for a non-anthropocentric approach (e.g. Bown, 2012, 2015; Roudavski & McCormack, 2016; Guckelsberger, Salge & Colton, 2017; McCormack, 2019). They agree that notions of creativity can be expanded beyond the time-scales and dimensions of normal human experience, behaviour and conventional thinking. Following an enactive account of cognition, we argue elsewhere (Guckelsberger, Salge & Colton, 2017) that there may be instances of creativity in artificial and biological systems that might not be recognised by people because of their embodiment and inherent bias towards human creativity. Yet, learning to recognise such ‘hidden creativity’ spurs the design of ‘creativity as it could be’ (Saunders & Gero, 2001b, p. 113), and can ultimately benefit people. McCormack (2019) agrees that studying CC from a non-anthropocentric angle can highlight ‘new possibilities for what constitutes creativity and creative behaviour’ (ibid., p. 328) with the potential of ‘bringing a more general theory to the fore’ (McCormack, 2019, p. 328; citing Roudavski and McCormack, 2016). As a starting point for a non-anthropocentric account, we propose to free the concepts of novelty and value in creativity from a bias shaped by our human experience. This
may allow us to investigate creativity from different viewpoints, while still including and explicitly encouraging the study of human creativity.

Based on the previous points, we formulate a working definition of CC that is more inclusive than existing accounts. We do so to tear down boundaries to related fields, and to reveal more relevant work in our systematic review:

Computational creativity (CC) is the explicit, multi-disciplinary study, both theoretical and applied, of creativity in any type of computational system. It considers creativity as an open-ended concept, at any level of complexity and from any viewpoint.

We acknowledge that the notion of CC is widely used to denote a specific research field and community that has substantially pushed related research over the last decades. With this working definition though, we promote CC as a general scientific endeavour, a lens of study that can be applied across various fields and communities. This allows us to resolve CC’s relationship with AI more generally and to state: whenever AI researchers explicitly theorise about or design a model of creativity, they engage in CC research.

In this definition, we write multi-disciplinary to stress that CC is being investigated in a distributed fashion from the perspectives of different disciplines. Still, we only identify research with CC if creativity is addressed explicitly\textsuperscript{10}. We write theoretical and applied, to emphasise both ends of studying CC without imposing any restrictions on particular methods. We limit CC to the study of creativity in computational systems to draw a clear distinction from creativity studies (cf. Sec. 4.1) and to emphasise the requirement for CC systems to take over some creative responsibility, rather than exclusively supporting their users’ creativity. To include related research in the widest range of AI sub-disciplines possible, we consider any type of computational system. We implicitly assume that creativity can be understood as specific instances of problem-solving, to ease comparisons with related work in AI. In contrast to other definitions (Sec. 4.2.1), we make the open-ended nature of the underlying creativity concept explicit. We consider work addressing any conceptualisation of creativity, and in particular creativity at any level of complexity. This includes big- to mini-c, but also other forms of creativity that can only be understood from different, potentially non-human viewpoints. This definition thus encompasses the continuum between the engineering and cognitive perspective (Sec. 4.2.1) expressed in existing definitions, and extends it further beyond its inherent anthropocentrism.

We next apply this and our working definition of IM (Sec. 2.2.3) in a systematic review of related work to answer this chapters’ research questions.

### Systematic Review

We are now sufficiently equipped to answer our first two research questions, ‘Why have IR and models of IM been used in CC?’ (RQ.3), and ‘RQApplicationsIMCC’ (RQ.4). To this end, we conduct a systematic review of existing theoretical and applied work using IR and models of IM in CC.

\textsuperscript{10} Although much of contemporary AI research touches on questions of creativity, we are thus reluctant to consider it CC, as long as this treatment is not explicit.
4.2.2.1 Method

In light of the contentious and ongoing definition of CC and IM, respectively, a challenge for this review has been to compromise between keeping the amount of considered related work manageable and being as inclusive and representative as possible. To this end, we have employed a two-stage process of (i) identifying and coarsely filtering candidate literature by their title and abstract and (ii) extracting qualifying work based on in-depth reviews.

For our first step (i), we systematically skimmed through publications at traditional venues for CC research and related conferences at the intersection of creativity and computational modelling: the International Joint Workshops on Computational Creativity (IJWCC) (2004-2008) as a part of the International Joint Conference on Artificial Intelligence (IJCAI), the European Conference on Artificial Intelligence (ECAI), and the AAAI Conference on Artificial Intelligence; the International Conference on Computational Creativity (ICCC) (2010-2019); the Computational Creativity Workshops of the Society for the Study of Artificial Intelligence and Simulation of Behaviour (AISB) (1999-2019); the Creativity & Cognition Conference (1993-2017); the International Conference on Computational Intelligence in Music, Sound, Art and Design (EvoMusArt) (2011-2019); the Genetic and Evolutionary Computation Conference (GECCO) (1999-2020); the Computational Creativity, Concept Invention, and General Intelligence Workshop (C3GI) (2012-2016); as well as the Workshops on Constructive Machine Learning (2016) and Machine Learning for Creativity and Design (2017, 2018) at the Conference on Neural Information Processing Systems (NeurIPS). We also sought relevant work in CC monographs and special journal issues. Using this initial selection, we traced references to additional, potentially relevant publications.

For our second step (ii), we have filtered the set of candidates based on detailed reviews and comparisons against our working definitions of CC (Sec. 4.2.1) and IM models (Sec. 2.2.3). Following our definition of CC, we have excluded AI publications that did not address creativity explicitly. Some of these cases were just marginally out of scope, e.g. Mahadevan’s (2018) proposal of ‘imagination machines’, which touches on many central CC themes, without explicitly relating imagination to creativity. We have not included work on computational game creativity, as this is treated separately in Ch. 5. Since the notion of IM has only been adopted in AI rather late, and because it might be used differently from our conception, we examined the fit of individual related work candidates to our diagnostics of IM models (cf. Sec. 2.2.3). While this has been challenging due to ambiguity or lack of detail in the model descriptions, it affords a more fine-grained comparison of existing approaches. We have found and included models that only partially qualify as IM, and thus only warrant some of the (yet to be presented) benefits for CC. Vice versa, we have excluded any work which does not model an agent with either physical or virtual embodiment as this would violate the most central diagnostic of agent-centricity, and serves as a prerequisite for some of the other diagnostics. Most notably, this applies to variations of what is commonly known as novelty (Lehman & Stanley, 2011), surprise (Liapis et al., 2013; Yannakakis & Liapis, 2016) and quality-diversity search (Pugh, Soros & Stanley, 2016). Furthermore, we have excluded agent-less generative models such as creative adversarial networks (Elgammal et al., 2017).
We have also omitted work in which a system’s dynamics are governed by other motivational mechanisms and IR has only a minor influence, such as the developmental model of early creative behaviour by Aguilar and Pérez y Pérez (2014, 2015, 2017). We finally only incorporate work that uses IR exclusively at some level of abstraction, as the joint optimisation of intrinsic and extrinsic reward obfuscates the benefits of the former over the latter.

Our final selection comprises 29 related work items spanning two decades, from as early as 1998 to 2018. We summarise the system details, creative domain, and the specifics of the used IM model in Tbl. 4.1. We elaborate on the detailed distinctions along the report of our findings in Sec. 4.2.2.2.

We answer RQ.3 with a typology\footnote{We exploit the key characteristic of typologies to ‘represent concepts rather than empirical cases’ (Smith, 2002, p. 381), with each dimension being ‘based on the notion of an ideal type’ (ibid., p. 381). The reasons to use IR and IM models can rarely be found individually or in their pure form in existing applications, but rather blended together for emergent benefits. Vice versa, related work can rarely be considered with respect to a single reason only.} of reasons to embrace models of IM in CC. On their own, these are (emergent) properties of both IR and intrinsically motivated behaviour, derived from our working definition of IM (Sec. 2.2.3). We determined their relevance for CC, thus trimming the set of candidate properties, in a three-stage process: Firstly, we identified concrete applications of IR and IM in related work. Secondly, we summarised these instances into abstract applications. Thirdly and most critically, we considered each abstract application a result of combining (i) such properties and (ii) theories of (computational) creativity, the latter being informed by the preceding Sec. 4.1.1, 4.1.2 and 4.2.1. These abstract applications also represent a typology, and we later use it to partially answer RQ.4. Both typologies only reflect what has been covered in existing work; we discuss how our contributions advance this state-of-the-art in Sec. 4.2.4.

4.2.2.2 Findings

Fig. 4.2 illustrates our findings. We have identified four properties of intrinsic reward (IR) (R.1-4), four properties of intrinsically motivated IM) behaviour (B.1-4) and two corollaries (C.1, C.2) that, in combination with 14 clusters of (computational) creativity theories, benefit 12 abstract applications of IR and IM in CC (A.1-12). We next provide a detailed account of the related work, structured under the abstract applications that the individual properties in the typology afford. By moving along the order of dependencies, we start with properties of IR and transition to properties of intrinsically motivated behaviour. We do not cover related work chronologically, but introduce it along with the abstract applications it serves most. For a summary of our findings, see Sec. 4.2.2.3.

Assessing p-creativity (A.1, A.2)

A first but insufficient reason to embrace IR in CC is that a specific instance of an IR function can be attributed different semantics (R.1). As a corollary, specific IR functions can capture variations of novelty and value (C.1). Based on the prominent role of these two concepts in the ‘standard definition’ (Runco & Jaeger, 2012) and in other conceptions of creativity, IR can be applied to assess...
### Table 4.1: Reviewed work using IR and IM in CC

<table>
<thead>
<tr>
<th>Authors</th>
<th>Year</th>
<th>Identifier</th>
<th>Domain</th>
<th>System Type</th>
<th>Study / System Details</th>
<th>Reward Properties</th>
</tr>
</thead>
<tbody>
<tr>
<td>Saunders and Gero</td>
<td>2001</td>
<td>Curious Design Agent</td>
<td>Autonomous Creativity, Co-Creativity, Scientific</td>
<td>A</td>
<td>S A S D V</td>
<td>Hedonic Novelty Max.</td>
</tr>
<tr>
<td>Saunders and Gero</td>
<td>2001</td>
<td>Curious Design Agent</td>
<td>Autonomous Creativity, Co-Creativity, Scientific</td>
<td>A</td>
<td>S A M D V</td>
<td>Hedonic Novelty Max.</td>
</tr>
<tr>
<td>Saunders and Gero</td>
<td>2001</td>
<td>Curious Design Agent</td>
<td>Autonomous Creativity, Co-Creativity, Scientific</td>
<td>A</td>
<td>E A M I V</td>
<td>Hedonic Novelty Max.</td>
</tr>
<tr>
<td>Cross</td>
<td>2006</td>
<td>Social Creation of Value in Evolution</td>
<td>Autonomous Creativity, Co-Creativity, Scientific</td>
<td>M</td>
<td>S T M D</td>
<td>Hedonic Novelty Max.</td>
</tr>
<tr>
<td>Merrick</td>
<td>2008</td>
<td>Curious Robots for Creative Play</td>
<td>Autonomous Creativity, Co-Creativity, Scientific</td>
<td>B</td>
<td>E A S D P</td>
<td>Hedonic Novelty Max.</td>
</tr>
<tr>
<td>Vigorito and Berto</td>
<td>2009</td>
<td>Intrinsically Motivated Creative Search</td>
<td>Autonomous Creativity, Co-Creativity, Scientific</td>
<td>S</td>
<td>S T S D</td>
<td>Competence (+)</td>
</tr>
<tr>
<td>Saunders</td>
<td>2010</td>
<td>Curious Design Assistants</td>
<td>Autonomous Creativity, Co-Creativity, Scientific</td>
<td>A</td>
<td>E A S X V</td>
<td>Hedonic Novelty Max.</td>
</tr>
<tr>
<td>Schmidhuber</td>
<td>2010</td>
<td>Formal Theory of Creativity, Fun, and IM</td>
<td>Autonomous Creativity, Co-Creativity, Scientific</td>
<td>V</td>
<td>S A S X</td>
<td>Novelty, Learning Progress Max.</td>
</tr>
<tr>
<td>Saunders et al.</td>
<td>2010</td>
<td>Curious Whispers</td>
<td>Autonomous Creativity, Co-Creativity, Scientific</td>
<td>M</td>
<td>S A M D P</td>
<td>Hedonic Novelty Max.</td>
</tr>
<tr>
<td>Saunders</td>
<td>2011</td>
<td>Curiosity-Driven Evolution of Language</td>
<td>Autonomous Creativity, Co-Creativity, Scientific</td>
<td>L</td>
<td>S A M D P</td>
<td>Hedonic Novelty Max.</td>
</tr>
<tr>
<td>Saunders</td>
<td>2012</td>
<td>Autonomous Creative Systems</td>
<td>Autonomous Creativity, Co-Creativity, Scientific</td>
<td>V</td>
<td>S T X D X</td>
<td>Hedonic Novelty Max.</td>
</tr>
<tr>
<td>Smith and Garnett</td>
<td>2012</td>
<td>RL &amp; Automated Musical Improvisation</td>
<td>Autonomous Creativity, Co-Creativity, Scientific</td>
<td>M</td>
<td>E A S D V</td>
<td>Hedonic Learning Progress Max.</td>
</tr>
<tr>
<td>Smith and Garnett</td>
<td>2012</td>
<td>Improvement With Hierarchical Neural Nets</td>
<td>Autonomous Creativity, Co-Creativity, Scientific</td>
<td>M</td>
<td>E A S D V</td>
<td>Hedonic Learning Progress Max.</td>
</tr>
<tr>
<td>Gemeinboeck and Saunders</td>
<td>2013</td>
<td>Zwischenraume</td>
<td>Autonomous Creativity, Co-Creativity, Scientific</td>
<td>B</td>
<td>A A M D P</td>
<td>Hedonic Novelty Max.</td>
</tr>
<tr>
<td>Saunders, Chee and Gemeinboeck</td>
<td>2013</td>
<td>Human Interaction With Curious Whispers</td>
<td>Autonomous Creativity, Co-Creativity, Scientific</td>
<td>M</td>
<td>S A M D P</td>
<td>Hedonic Novelty Max.</td>
</tr>
<tr>
<td>Grace and Maher</td>
<td>2015</td>
<td>Specific Curiosity &amp; Transformational Creativity</td>
<td>Autonomous Creativity, Co-Creativity, Scientific</td>
<td>C</td>
<td>S T S D</td>
<td>Surprise, Specific Surprise Max.</td>
</tr>
<tr>
<td>Linkola, Takala and Toivonen</td>
<td>2016</td>
<td>Novelty-Seeking Creative Multi-Agent Systems</td>
<td>Autonomous Creativity, Co-Creativity, Scientific</td>
<td>A</td>
<td>E A M D V</td>
<td>Novelty Max.</td>
</tr>
<tr>
<td>Wiggins</td>
<td>2018</td>
<td>Information Dynamics of Thinking</td>
<td>Autonomous Creativity, Co-Creativity, Scientific</td>
<td>M</td>
<td>S T S D</td>
<td>Model Complexity, Novelty Min. Max. (+)</td>
</tr>
</tbody>
</table>

We have omitted similar work by Saunders and Gero (2001c, 2002) and Saunders (2006) as it does not express further reasons to embrace IR and IM models in CC. Similarly, this summarises earlier related work (Schmidhuber, 1997, 2006, 2007) and is later on extended slightly in (Schmidhuber, 2012).
Figure 4.2: Typologies of the benefits and applications of intrinsic reward (IR) and intrinsically motivated (IM) behaviour in CC, derived from existing work. We conceive (abstract) applications to arise from combining: (i) (emergent) properties of IR, intrinsically motivated behaviour and corollaries, following from our IM definition; and (ii) theories of (computational) creativity.
the creativity of (partial) artefacts (A.1). However, similar interpretations exist for extrinsic reward, and the use of IM in existing CC studies can thus only be understood in conjunction with a second reason: based on the defining diagnostics of agent-centricity and embodiment universality, IR is subjective and sensitive to an agent’s embodiment and situatedness (R.2). IR thus allows us to model p-creativity (A.2, cf. Boden, 2003; and Sec. 4.1.1) as a subjective, embodied and situated assessment of novelty and value. In related work, subjective novelty is often hidden in the notions of unexpectedness and surprise; while novelty is commonly formalised as the degree to which an outcome is different from prior outcomes experienced by an agent, unexpectedness measures how much an outcome deviates from expected outcomes predicted by an agent-subjective model (Grace & Maher, 2015). Surprise belongs to a separate category, as it denotes an affective response to unexpectedness (Pearce & Wiggins, 2012; Grace & Maher, 2015).

Macedo and Cardoso (2001a) assume that a creative product must not merely be novel, but bear unexpected novelty. Since such unexpectedness leads to the affective state of surprise, they concede that surprise must play an important role in the evaluation of creativity in both artistic and scientific products. In a prototype system, they rank an existing set of simple, two-dimensional architectural sketches with respect to the surprise computed by an individual agent. While considering it an affective state, surprise is calculated directly as the unexpectedness of perceiving a new object, given the agent’s experience of previously seen objects. This is formalised by averaging and then complementing the conditional probabilities of perceiving each of the object’s individual components, given the previously memorised objects.

In a theoretical contribution, McGregor (2007) roots agent-centric and embodiment-sensitive measures of novelty at the very basis of the CC agenda. He argues that assessing novelty with an ‘objective’ or ‘impersonal’ distance metric such as the information distance (Bennett et al., 1998) between artefacts ‘would be at fundamental odds with the goals and methods of computational creativity’ (McGregor, 2007, p. 111) as it would rule out the existence of a compact algorithm capable of generating maximum novelty with respect to known examples. Referencing the intrinsic novelty functions by Saunders (2001) and Schmidhuber (2006), he highlights the need for a ‘personal’, ‘perceptual’, observer-relative account of novelty under which ‘endless apparent novelty could be generated by a compact program by exploiting the limitations of the perceiver’s ability to detect patterns’ (McGregor, 2007, p. 111).

INCREASING CREATIVE AUTONOMY (A.3)
One fundamental goal of CC researchers is to endow artificial systems with creative responsibility (Colton & Wiggins, 2012) or creative autonomy (Jennings, 2010). Without such autonomy, a system could not genuinely introduce novelty in the creation of artefacts, and would thus become a victim of Lovelace’s famous objection to originality in Babbage’s Analytical Engine: ‘The Analytical Engine has no pretensions to originate anything. It can do whatever we know how to order it to perform’ (Menabrea, 1842, p. 722). The need for

---

12 In the latter case, novelty is thus assessed relative to non-materialised predictions that are based on an agent’s past experiences, rather than directly with respect to these experiences.
autonomy is also justified empirically: autonomy is a commonly observed character trait of creative people (Sheldon, 1995; Davis, 1999; and Sec. 4.1.2), and has been highlighted as a central criterion for people to attribute creativity to a machine (Mumford & Ventura, 2015). It follows from the agent-centricity and embodiment universality diagnostics that a purely intrinsically motivated agent neither can, nor has to follow instructions formulated at runtime in the form of extrinsic reward; instead, they generate their own subjective goals in response to their embodiment and situatedness (R.2). Together with the potential for IR to model novelty and value (C.1), IM allows us to increase the level of creative autonomy in CC systems (A.3).

Saunders (2012) summarises various meanings of autonomy into the requirement for ‘self-law making, or self-governing’ (ibid., p. 219). He argues that Boden’s (2003) theory of creativity is insufficient for the development of such autonomous creative systems because it misses an account of motivation. Based on a review of applied related work13, he argues that models of IM can fill this gap, thus allowing for a system’s self-governance and reducing dependencies on designer-imposed, fixed rules.

The importance of IM for creative autonomy has also been addressed in applied work. Gemeinboeck and Saunders (2013) stress that IM-induced creative agency is not predetermined ‘but evolves based on what happens in the environment [that agents] examine and manipulate. As the agents’ embodiment evolves based on its interaction with the environment, the robots’ creative agency affects processes out of which it itself is emergent’ (ibid., p. 218). Saunders et al. (2010) highlight that as consequence, the interaction dynamics between people and CC systems, here as robots, are transformed: people enter the interaction as equals, rather than superiors, being denied the ability to ‘dictate’ what the robots should do; Saunders and Gemeinboeck (2014) note that ‘rather than being invited to control the course of events, the audience is implicated in the material interventions’ (ibid., p. 3).

**Substituting extrinsic reward (A.4)**
Most of the work from the previous paragraph has been motivated from a cognitive (Saunders et al., 2010) or artistic (Gemeinboeck & Saunders, 2013; Saunders & Gemeinboeck, 2014) perspective, where limiting people’s control over the behaviour of CC systems has been an explicit goal. However, reducing the need for control and increasing creative autonomy becomes more complicated from an engineering perspective: CC researchers typically want to leverage autonomous or co-creative systems for the benefit of people, e.g. by solving a particular task or generating an artefact which they appreciate. This produces a conflict, in that explicitly communicating such external goals to a system diminishes their creative autonomy. We face similar challenges also through the cognitive perspective: If we want to learn more about human creativity, we must be able to explain why intrinsically motivated people can score highly on extrinsically defined tasks as present in e.g. a workplace creativity setting (cf. Sec. 4.1.2).

Both issues are addressed by the third reason to embrace IM in CC: IR can align with extrinsic reward (R.4), either produced by another computational

---

13 We relate to this work in detail under the separate abstract applications A.6-A.11.
model, or by an organism. An intrinsically motivated agent thus has the potential to relate the (partial) outcomes of their decision-making to the extrinsic but unknown goals of other agents. Given that IR can align with extrinsic reward, and that human-intrinsic reward qualifies as such agent-extrinsic reward, it crucially follows that some instances of computational IR can correlate with human IR (C.2). McGregor (2007) highlights the diversity that e.g. personal (and embodied) measures of novelty afford: the same sequence of outcomes could be considered highly novel by one and highly similar by another individual. This diversity presents an opportunity to solve the control dilemma: by appropriately modelling and tuning the components that shape IR, i.e. an agent’s specific embodiment and the parameters of their reward function, we can control the alignment of IR. IR has consequently been used to substitute extrinsic reward\(^{14}\) (A.4).

APPROXIMATING HUMAN AESTHETIC JUDGEMENT (A.5)

One specific use-case for such substitution is given when extrinsic reward is either sparse or unavailable. A notorious example is human aesthetic preference: ‘Our aesthetic values are difficult to recognize, more difficult to put into words, and even more difficult to state really clearly’ (Boden, 1990/2003, p. 10). But the ability to predict a user’s aesthetic response to an artefact, either to produce an artefact which matches their aesthetic expectations or to subvert them, is an essential requirement for the artistically creative type of system that has dominated CC for a long time. Unsurprisingly, some of the earliest studies of IR in CC have focussed on exploiting IR to approximate human aesthetic judgement (A.5). Rather than focussing on human-extrinsic, e.g. culturally shaped, aesthetics, they draw on the work of e.g. Berlyne (1971) and Williams (1996), who relate novelty and aesthetic preference. Existing work thus approximates human aesthetic preference by aligning agent-intrinsic with human-intrinsic novelty, exploiting corollaries C.1 and C.2.

Schmidhuber’s (1997) Formal Theory of Beauty is an important predecessor to these studies. It rests on the assumption that any human observer tries to represent input data in terms of what is familiar, i.e. they compress new stimuli based on experience. The maximum achievable compression then corresponds to a stimulus’ novelty or conversely, familiarity. Schmidhuber formalises this novelty via Kolmogorov complexity (Cover & Thomas, 2006, p. 144 ff.), i.e. the length of the shortest algorithm capable of computing the stimulus. Inspired by formal rules in art and aesthetics, he proposes that the most beautiful stimuli are the ones that can be compressed most, given the individual’s current compression scheme. The latter is assumed to be shaped by experience, which would explain individual differences in the perception of beauty. Schmidhuber thus argues for subjective novelty to work as a proxy to the human perception of beauty, and equates maximum beauty with maximum compression or familiarity. Crucially though, he deems this formal account of beauty only a pre-stage to aesthetic reward\(^{15}\) (Schmidhuber,

\(^{14}\) This abstract application does not follow straight from related work, but we yet include it to bridge smoothly to A.5 and to motivate our contributions later on.

\(^{15}\) The Kolmogorov complexity is furthermore generally intractable, rendering it unavailable to an agent’s decision-making. We consequently omit this contribution from Tbl. 4.1.
and does not test the hypothesised relationship of compression and beauty by sufficiently strong empirical means.

Apparently unaware of Schmidhuber’s work, Peters (1998) proposes such an IR to endow a machine with the means for aesthetic evaluation, through which it could filter either algorithmically generated or ‘found’ images. He takes inspiration from the theoretical work of Williams (1996), who, by drawing on Leyton (1992) and by reference to cognitive dissonance (cf. Sec. 2.1.2), argues that the human aesthetic response is proportional to the amount of *explicable surprise*, i.e. the amount of change in an individual’s beliefs caused by the perception of an unexpected stimulus. The stimulus must be *surprising*\(^\text{16}\), i.e. unexpected given the agent’s present beliefs shaping their predictions, to provide a scope for explanation. *Explicability* means that the agent must be capable of changing their beliefs to confirm with the new perception. Peters formalises surprise as a prediction error of individual pixel values in a robot’s sensor. He then constructs an explicability reward by stringing together several of these surprise functions, thus rewarding a change in prediction error. Peters thus presents an alternative formalisation of Schmidhuber’s (1991) *learning progress*, motivated as an aesthetic reward.

Schmidhuber recognises the role of novelty or surprise in the modelling of aesthetic reward later as part of his *Formal Theory of Creativity* (Schmidhuber, 2006; Schmidhuber, 2012; in Tbl. 4.1 comprised in the 2010 survey article). Similar to Williams (1996), he considers something to remain aesthetically rewarding or *interesting* for an observer as long as the observer makes progress on explaining it further through adaptation of a predictive model. He thus formalises aesthetic reward as the first derivative of subjective beauty. Based on his hypothesised formalisation of beauty (Schmidhuber, 1997), this can be calculated as the *difference* in the data’s *Kolmogorov complexity* before and after learning. At a certain time, something may be beautiful but not aesthetically rewarding, as the agent cannot compress it further based on their present experience (Schmidhuber, 2010). Since Schmidhuber (1991b) proposes prediction error as a tractable alternative to Kolmogorov complexity, and formalises learning progress as the difference of prediction error, his and Peters’ (1998) account of aesthetic reward are closely related.

Saunders (2009) probes the correlation between IR and human aesthetic preference in an applied study. He extends earlier proposals of supporting a human artist in their creative activity by letting an artificial agent pre-select ‘interesting’ instances of computer-generated artefacts, thus guiding the artist’s attention and preventing information overload (Saunders & Gero, 2001a, 2001d). The intrinsic interestingness reward used is also adopted from earlier work (Saunders & Gero, 2001a), and, similarly to Peters (1998), is also a function of novelty; Saunders and Gero’s (2001) formalisation however draws on Berlyne’s (1971) popular theory of human aesthetic preference, and could thus be argued to have the stronger grounding. Supported by empirical data, Berlyne (ibid.) models the relationship between arousal, e.g. in the form of novelty, and ‘hedonic value’ such as interest, by means of the Wundt curve: a non-linear, inverted U-shaped function conceived by Wilhelm Maximilian

\(^{16}\) Williams (1996) understands surprise in strict information-theoretic terms as the *self-information* (cf. Appx. C) of an event, e.g. the perception of a specific sensor state.
Wundt (1874). This ‘hedonic function’ outputs high interest for stimuli that are neither too familiar nor too novel to previous experiences. Saunders and Gero (2001a) quantify the underlying novelty with a self-organising map (Kohonen, 1995) which learns to categorise high-dimensional sensory inputs in a lower-dimensional space. Each neuron in the map corresponds to one category, and an input’s novelty corresponds to the inverted activation of the best-matching neuron, indicating the typicality of the input for that category. Interest or hedonic novelty reward\(^\text{17}\) is then computed by passing the novelty value through a parametrised hedonic function. In the 2009 applied study, a curious design assistant ranks 9 generative artworks produced in each generation of an evolutionary algorithm with respect to the hedonic novelty reward they afford, measured on their experience of artefacts generated in previous iterations. Over 3 trials with 50 iterations each, the agent’s top rank matches the user’s preference with 30-50% accuracy. Taking only the first three ranks into account yields a higher accuracy of 60-72%.

The previous applications leverage (emergent) properties of IR only. We next discuss how IR has been used in decision-making to model the creative process. Related applications hereby exploit (emergent) properties of intrinsically motivated behaviour, arising from combining specific IRs with specific action-selection functions.

**Modelling Exploratory Creativity (a.6, a.7)**

Boden (1990/2003) conceives creativity as the identification of surprising and potentially valuable concepts in a conceptual space. She hereby distinguishes between three mechanisms of creativity, from which exploratory and transformational creativity have been modelled with IM. Exploratory creativity consists of discovering novel and potentially (but not necessarily) valuable\(^\text{18}\) concepts within a known conceptual space. This resonates well with creativity studies’ emphasis on the central role of IM in human creativity (Amabile, 2012), motivating them to collect novel information (Liu et al., 2016) and envisage different possibilities (Kieran, 2014). Based on his creative systems framework, Wiggins (2006a, 2006b) formalises Boden’s account as search in such a space. For such search to qualify as exploratory creativity, an agent must be able to access new concepts in the conceptual space by following a fixed set of traversal rules, which may utilise heuristics evaluated on (partial) concepts (Wiggins, 2006b). Crucially, specific instances of IM models as combinations of a specific IR and action-selection function allow for such exploratory behaviour (B.1). This is usually contingent on using IR that can be interpreted as novelty and value (C.1). Mapped to Wiggins’ framework, the IR function can be considered a heuristic used by a specific, fixed action-selection strategy.

---

\(^{17}\) Since the hedonic function could be parametrised to yield the highest interest for stimuli that stipulate but do not overburden the learning capacity of a specific agent, the resulting reward can be interpreted as a hard-coded approximation of Schmidhuber’s (1991) learning-progress.

\(^{18}\) In her popular book, Boden (1990/2003) is not very clear about the importance of value for exploratory creativity: she stresses the discovery of novel concepts to learn about the possibilities of a conceptual space, but elaborates little on the notion of value. However, she elsewhere relates to value by noting that a ‘main reason why most current AI-models of creativity attempt only exploration, not transformation, is that if the space is transformed, then the resulting structures may not have any interest or value. Such ideas are novel, certainly, but not creative’ (Boden, 1998, p. 354, emphasis added).
such as maximising (expected) IR. Researchers have thus embraced IM to model exploratory creativity (A.6).

Although the following examples can be considered cases of exploratory creativity in a post-hoc analysis, most have not been motivated as a study of this specific mechanism, but of the creative process more generally. In many cases, the IR function is chosen specifically to approximate human aesthetic judgement (A.5) as a basis for producing artefacts that appeal to people (A.7). They hereby exploit the emergent property of intrinsically motivated behaviour to yield task performance in the presence of sparse extrinsic reward or its total absence (B.2). This again rests on the potential of IR to align with extrinsic reward (R.4), and, for the application to aesthetics, on its correlation with human IR (C.2).

Schmidhuber (1997) again precedes related work by proposing the generation of low-complexity art as an iterative drawing process in which a drawing’s Kolmogorov complexity is minimised, which he claims leads to the maximisation of beauty. We consider this only a predecessor to related compression-based CC work because (i) calculating the Kolmogorov complexity is generally intractable, (ii) the latter is observer-dependent but the ‘typical observer’ is not formally accounted for, and (iii) the process still relies on people to judge drawings to ‘look right’ (ibid., p. 97). Schmidhuber concludes: ‘No universal algorithm for generating low-complexity art is known. At the moment, a human artist is still required’ (ibid., p. 102).

Macedo and Cardoso (2001a) propose to use the surprise reward introduced earlier as a means to advance an architectural sketch. They appeal to exploratory creativity by noting that ‘surprise plays an important role to make the process of producing products (...) divergent’ (ibid., p. 3). Actions hereby map to adding a new component, e.g. a window, to an existing sketch, and their action-selection function greedily chooses the component that would yield the combined sketch with the highest expected surprise. They demonstrate the calculation of expected surprise for different combinations, but do not provide an example for the time-expanded creative process, hereby circumventing the question of when a design can be considered finished.

Saunders and Gero (2001a) highlight three applications of artificial curiosity to support design: (i) to search and explore unfamiliar design spaces and gain a better understanding of a non-routine task, (ii) to guide problem-solving to find interesting design solutions, and, on a meta-level, (iii) to guide problem-finding to discover interesting design problems. They focus on the first case of value-less exploratory creativity, and formalise a curious design agent that is rewarded by a design’s hedonic novelty, as introduced earlier. Saunders and Gero demonstrate their model on the generative space of two-dimensional spirograph patterns, where actions correspond to a change in the spirograph’s wheel ratios, and the action-selection function yields fewer changes when interest is high. A simulation over 200 time steps shows patterns of novelty peaks that tail off as soon as the self-organising map underlying the reward calculation has learned a new ‘design category’. The initially low novelty leads to high changes to the spirograph wheels in the beginning, thus producing a diverse set of categories for complex patterns early on. Saunders and Gero conclude that curiosity provides a ‘general-purpose, knowledge-lean heuristics to guide the search for potentially interesting, and possibly even creative, design’ (ibid., p. 350). Cross (2006) later demonstrates the robustness
of the *curious design agent* with respect to extreme parameterisations of the hedonic function underlying the reward calculation. The agent keeps exploring even for parametrisations that express a preference for high novelty or familiarity, respectively. Saunders (2009) later motivates the use of curious design agents to support human creative activity, rather than design alone. He links curiosity to creativity by ‘the exploratory behaviour typical of the early stages of a creative activities [sic] as a user learns about the possibilities within a space’ (ibid., p. 3), allowing to ‘relieve the uncertainty that accompanies an incomplete understanding of the conceptual space’ (ibid., p. 2). Boden (1990/2003, p. 58f.) has highlighted the usefulness of such state coverage as preparation for goal-directed creativity, the first stage of creative thinking put forward by Poincaré (1913) and Hadamard (1954). Saunders (2009) proposes to use *curious design assistants* as a form of curious design agents to ‘autonomously exploring a design space using a generative design system and then prepare a report on potentially interesting solutions at the end of the exploration’ (ibid., p. 3). However, a demonstration is missing.

Smith and Garnett (2012b) use a *hedonic learning progress* reward to model an agent capable of improvising and analysing music that is valued by people. They motivate their choice by summarising that ‘self-motivated reinforcement learning models present new possibilities in computational creativity, conceptually mimicking human learning to enable automated discovery of interesting or surprising patterns’ (ibid., p. 223). In an effort to align the agent’s perception of novelty with a person’s perception (cf. McGregor, 2007), they extract a number of hand-crafted musical features from the raw musical score. Similarly to Saunders and Gero (2001a), they first calculate learning progress on the weight changes of an adaptive resonance theory network (Carpenter, Grossberg & Rosen, 1991), which learns to categorise the feature vectors. This quantity is then transformed through a hedonic function to capture ‘yet unexplained but easily learnable regularity’ (Smith & Garnett, 2012b, p. 224). The agent’s actions are given by changes in the chromatic pitch, and action-selection is driven by greedily maximising the *hedonic learning progress* reward. Three experiments show that depending on the network’s learning rate, the agent explores all pitch classes or finds more interest in repetition. Biasing the network by pre-training it on Bach’s cello suites yields more complex structure in the produced score, with some examples bearing similarity with the input suites. Smith and Garnett (2012a) try to increase the system’s sophistication by stacking several networks, thus abstracting the manually engineered features on the first layer further into more complex ones. The *hedonic learning progress* is now computed by contrasting the expansion of the networks with the amount of change in their nodes, thus quantifying maximum interest between boredom and chaos. Experiments across different variations of the hierarchical model yields ‘musical material which revealed pattern repetition and manipulation in [a] non-obvious, yet intelligible fashion’ (ibid., p. 66).

---

19 Coverage of the entire conceptual space should be considered a specific, asymptotic case of exploratory creativity. It is generally not required for behaviour to qualify as exploratory creativity, and is crucially not supported by every model of IM that warrants the latter.
MODELLING TRANSFORMATIONAL CREATIVITY (A.7, A.8)

The restriction of exploratory creativity to a fixed conceptual space and search strategy limits the identifiable concepts to those that are conceivable within that space, and that can be reached with the given traversal strategy. Transformational creativity widens this scope by inducing a change of rules (Boden, 1990/2003). Wiggins’ (2006) search-based framework enables the distinguishing of R-transformation as changing the rules that define the very search space, and T-transformation as changing the traversal rules to navigate it (ibid.). R- and T-transformational creativity can thus be understood as exploratory creativity on a second-order representation (ibid.) of possible search spaces, and on the space of possible traversal strategies, respectively. Since IM models can yield exploratory behaviour (B.1), researchers have leveraged them to model heuristic transformational creativity (A.8). Again, much related work exploits this mechanism to produce artefacts that appeal to people (A.7). We assume that a system’s traversal strategy can be understood as an action policy, and that any decision-making framework that allows for the on-line adaptation of a policy necessarily has the potential for T-transformational creativity. We thus include any related work that uses IR in RL under this heading.

Peters (1998) identifies the exploration of a problem space as an essential component of productive thinking (Wertheimer, 1959), a Gestalt psychological concept which denotes the ‘solving of problems in a manner that is significantly new’ (Peters, 1998, p. 836); it involves ‘creating rather than recalling a solution’ (ibid., p. 836) and is closely related to McGraw’s (1978) notion of heuristic solutions (Sec. 4.2.1). Peters argues that exploration can help an individual to navigate a problem space while being unable to ‘estimate how close they are to a solution’ (ibid., p. 836). He proposes to drive such exploration by using his earlier introduced formalisation of surprise as an RL reward to be maximised, which allows for T-transformational creativity. Unfortunately, his proposal remains theoretical.

In his so-called Formal Theory of Creativity, Schmidhuber (2006 and 2012, in Tbl. 4.1 comprised in a 2010 survey article) not only introduces compression- or learning progress as an aesthetic reward, but proposes it as the foundation of any form of creative activity. He argues that creative people maximise the same reward at various stages of their lives, and across different creative domains such as jokes, music, art and scientific discovery. He moreover suggests that, depending on whether actions map to the creation or perception of art, we get creativity in production and aesthetics in reception, respectively: ‘When not occupied with optimizing external reward, artists and observers of art are just following their compression progress drive’ (ibid., p. 243). Schmidhuber understands scientific creativity as theory formation, i.e. the narrowing down of hypotheses to those that explain the observed phenomena best, thus yielding compression breakthroughs. While he argues for his account of beauty based on work in formal aesthetics, a theoretical or empirical basis for his claims on creativity is absent. None of the referenced agents using learning progress as a RL reward demonstrate the discussed scientific or artistic creativity, and we hence consider this only a theoretical contribution.

This is a preliminary assumption, as possible mappings between the creative systems framework and other decision-making frameworks are subject of active research. We recently contributed to this agenda in (Linkola, Guckelsberger & Kantosalo, 2020).
Colin et al. (2016) propose a theoretical mapping between psychological theories on *insight-driven problem-solving* as the ‘fast understanding of original, illuminating solutions to problems’ (ibid., p. 198), intrinsically motivated hierarchical RL, and *exploratory* and *transformational creativity* as formalised in the *creative systems framework* (Wiggins, 2006a, 2006b). They focus on two challenges of modelling insight-driven problem solving: to (i) optimise not only extrinsic reward, defining the problem to be solved, but also intrinsic novelty; while (ii) managing the complexity of a potentially enormous search space. As a solution, they propose to use a variety of well-established IRs for learning *options* (Sutton, Precup & Singh, 1999) in hierarchical RL. Options are ‘temporally extended, closed-loop courses of action’ (Colin et al., 2016, p. 201), and each option is given by a *policy* that is initiated and terminates in a certain state set, and a state *representation* that regulates which state features matter for the policy. They argue that options as temporal abstractions allow ‘for exploring the state space at multiple granularity sizes’, making it possible ‘to reach otherwise unattainable sections’ (ibid., p. 201). They relate this formalisation of insight-driven problem-solving to creativity by mapping the conceptual space in Wiggins’ framework to a problem-space, and each concept to an option. Consequently, the second-order representation of possible conceptual spaces becomes the space of possible options, each with a different representation of the problem space and an initial policy to navigate it. They conclude that under this mapping, *exploratory creativity* corresponds to the learning and following of a policy defined by a specific option working on a fixed state representation. Furthermore, they identify *transformational- as exploratory creativity* on the second-order representation in the switching between options, learned from experience. In their theory, both processes are driven by IR alone, or in combination with an extrinsic reward.

Grace and Maher (2015) theorise how different intrinsic surprise rewards can facilitate *creative intention* in problem-solving, accompanied by different forms of *transformational creativity*. They argue that ‘intentions are not created de novo, but (...) arise from a drive to explore what the system has observed but not understood’ (ibid., p. 261), and distinguish such *specific curiosity* from *diversive curiosity* as commonly used in related work. They formalise curiosity via an extension of the *creative systems framework* (Wiggins, 2006a, 2006b) and creative intention by distinguishing two types of search: In basic problem-solving, an agent seeks to maximise both value and surprise, i.e. they complement goal-optimisation with *diversive curiosity*. When experiencing a particularly surprising concept, the agent transitions to a second search mode in which they act towards maximising *specific surprise*, rewarded by concepts that violate the same expectations that triggered the search mode transition. *Intentional creativity*, according to Grace and Maher, then maps to the identification of specifically surprising concepts. Similar to Wiggins (2006a, 2006b), they understand *R-transformational creativity* as an agent’s update of their description of probable concepts in response to experiencing an inexplicable or unexpected concept. In addition, they identify *T-transformational creativity* in the ‘intentional’ switching to specific curiosity search. They argue that such mode-switching can be caused by exposure to concepts from an inspiring

---

21 They hereby miss that learning and thus adapting a policy can qualify as T-transformation.
set or to other agents, and by actively creating concepts that are expected to trigger a specific curiosity episode.

We chose these examples for their focus on exploratory and transformational creativity. More generally, all related work in Tbl. 4.1 that use IM in a decision-making process (e.g. Maher, Merrick & Macindoe, 2005; Merrick, Saunders & Maher, 2007; Merrick, 2008a; Vigorito & Barto, 2008b; Wiggins, 2018) exhibit the property (B.1) and thus realise at least exploratory creativity.

MODELLING SYSTEMS THEORIES OF CREATIVITY (A.9, A.10)
The previous examples use IM to drive the exploratory and transformational creativity of individual agents. While the goal is often to appeal to human aesthetics, the human audience remains a passive recipient of artefacts produced by a sole creative agent. This contrasts with systems theories of creativity (cf. Vygotsky, 1971; Csikszentmihalyi, 1988; and Sec. 4.1.1), stressing that creativity does not happen in a vacuum but results from the interaction of individuals in a larger societal and historical context. Very early on, IM has been used to drive individual agent behaviour in systems theories of creativity (A.9), by exploiting the subjective nature of IR and its sensitivity to different agent embodiments (R.2), as well as its possible interpretations with respect to the core components of creativity (C.1). As a closely related application (A.10), a systems model enables us to let h-creativity emerge from p-creativity (Boden, 2003; and Sec. 4.1.1). Existing work exploits the same properties to study how h-creativity is influenced by variations in individual agent embodiments.

Many examples of related work do not aim at producing artefacts that are cherished by human users, but embrace a cognitive perspective to shed more light on the functional underpinnings of (human) creativity. Macedo and Cardoso (2001b) e.g. focus on the central role of surprise in creativity, and investigate the relation between the level of surprise used in the production of an artefact by an author agent, and the surprise elicited in several critic agents. The author uses the previously introduced surprise reward (cf. Macedo & Cardoso, 2001a) to create an artefact through the incremental addition of maximally surprising components, eventually putting their creation into the environment for assessment by the critics. The critics are similarly driven by surprise but their actions are movements, bringing them closer to more surprising artefacts. Through experimentation with the design of architectural sketches, Macedo and Cardoso find that the higher the surprise in authoring the sketches, the higher the critics’ surprise. Counterintuitively, their reward function yields higher surprise for critics with large than for those with small memory. While this can be considered a systems model, it is limited by the absence of a feedback loop from the critics back to the author.

Saunders and Gero (2001d) are arguably first to formalise a closed-loop systems model of creativity. The goal of their Digital Clockwork Muse is to ‘explore the role that an individual’s search for novelty plays in socially situated creative systems’ (Saunders & Gero, 2002, p. 82). They specifically set out to probe several hypotheses by Martindale (1990), who highlights the maximisation of novelty as a key motivation of individuals in creative societies. He postulates that the complexity of a new style developed by novelty-seeking individuals must necessarily increase over time to satisfy an increasing global demand for novelty. Individuals thus obey what he calls
a ‘law of novelty’, punishing repetition and enforcing progress. Saunders and Gero model a society of agents that each maximise the *hedonic novelty* or *interestingness* (Saunders & Gero, 2001a) of individually evolved artworks. In contrast to Macedo and Cardoso (2001b), they combine the roles of author and critic. Each agent can perceive their own creation and those of other agents. If an individual deems their artwork interesting enough, they send it to other agents for review, or evolve it further if not. If an individual deems another agent’s artwork more interesting than theirs, they use it as a replacement for their own work to evolve further. The receiving agent can also push this artwork into a global repository, for future agent generations to begin their search with artworks that were once deemed creative. Their simulations support Martindale’s (1990) hypotheses and yield his *law of novelty*: Searching for higher hedonic novelty produces more complex artworks, and agents with similar preference for novelty are deemed creative by their peers, while others are isolated, eventually leading to the formation of cliques.

The *Digital Clockwork Muse* (Saunders & Gero, 2001d) instantiates what Saunders and Gero (2001b, 2001c, 2002) later introduce as *Artificial Creativity*, a framework for the study of creative societies (in Tbl. 4.1 comprised in Saunders and Gero (2001b)). They adopt the *cognitive perspective* for ‘the comparative study of creativity as it is found in human societies against creativity as it can be computationally modelled in artificial societies of agents’ (Saunders & Gero, 2002, p. 113). Their framework is strongly inspired by Csikszentmihalyi’s (1988) *domain-individual-field-interaction* systems model of creativity. Csikszentmihalyi considers creativity as emergent from the *interaction* between *individual* creators, a larger *field* of agents selecting contributions worth preserving, and a *domain* as a repository of artefacts held by a culture for future generations to draw from. This model has been adopted in Liu’s (2000) *dual generate-and-test* model of social-cultural creativity in the form of a generate-and-test loop at the level of society, complemented by a loop at the level of individuals. While Liu defines the socio-cultural evaluation explicitly, Saunders and Gero extend the framework to make it emerge from the evaluation of individual agents, driven by a model of curiosity. This modification turns the *Artificial Creativity* framework into a closed system (Saunders & Gero, 2001c) where no agent can dictate the behaviour of other agents and crucially, no rules dictate global behaviour (Saunders & Gero, 2001b). The emergence of socio-cultural evaluations from IM allows to investigate how a creative process results from a specific social situatedness of agents with potentially individual differences. The *Artificial Creativity* framework represents the blueprint for much later work on systems models of creativity.

Saunders and Gero (2004) later switch to the *engineering perspective* and use a crowd of intrinsically motivated agents to solve the task of designing exhibition layouts that avoid blockages and afford visitors exploration and learning. They approximate the behaviour of human visitors (cf. A.5) by simulating the movement of agents that greedily maximise *hedonic novelty* (Saunders & Gero, 2001a) in a virtual exhibition. At the end of each simulation, they assess how evenly the various categories of the self-organising map underlying the reward calculation have been covered. Since an even coverage indicates that the agents have seen a wide range of paintings, it aligns with the extrinsic goal of a good gallery layout. Saunders and Gero demonstrate this successful
alignment via simulations on pre-made exhibition layouts. The overarching system here cannot be considered an embodied agent, and does not use IR directly to drive its actions. It however still leverages information of intrinsically motivated, (virtually) embodied agents within, and we consequently consider this an indirect use of (intrinsic) reward.

Bown (2006) adopts Saunders and Gero’s (2001) Artificial Creativity framework (ibid.) to investigate the development of human artistic creativity from an evolutionary viewpoint, and exemplified through music. He proposes that music could have emerged from the basic effects of social value creation, arguing that ‘intraspecies competition is as much a driver for evolution as adaptation to an external environment’ (Bown, 2006, p. 1), and that especially ‘artefacts resulting from [artistic] acts have no value other than that offered by other individuals’ (ibid., p. 1), potentially raising their creator’s social status. He furthers this argument through the lens of the Artificial Creativity framework, equating social value with the hedonic novelty awarded between the agents through interaction in the field. Based on the underlying model of IM, he hypothesises how the society and the individual curious agents would change if they were subjected to evolutionary pressure for social value. He predicts that the agents will over time develop more sophisticated perceptual mechanisms and become more discriminatory. He also anticipates that individual preferences for novelty formalised through different parametrisations of the hedonic function, as well as social structures in the form of cliques enforcing a certain style, eventually evolve into a stable state.

Bown and Wiggins (2009) complement this theoretical contribution with an applied study, in which they maximise a simplified version of Saunders and Gero’s (2001) hedonic novelty reward to drive the individual development of musical style in an agent society. In this cultural dynamical system, each agent curiously explores new styles, and presents the style with the highest hedonic novelty to their neighbours. These in turn reward the agent with status proportionally to the hedonic novelty of the performed style compared to their own. Higher status provides privileged access to energy, which allows for this mechanism of musical ‘enchantment’ to indirectly increase individual fitness. Via several simulation experiments, Bown and Wiggins support Bown’s (2006) earlier claims and highlight that the evolution of human musical behaviour could have begun with maladaptive cultural dynamics that became reinforced through an overarching evolutionary process.

Saunders et al. (2010) later instantiate an Artificial Creativity system in the physical domain in the form of Curious Whispers, a society of mobile robots generating simple tunes and listening to the tunes produced by others, including people. Following Pickering (2005), Saunders et al. (2010) note that ‘creativity cannot be properly understood, or modelled, without an account of how it emerges from the encounter between the world and intrinsically active, exploratory and productively playful agents’ (ibid., p. 100). By embracing physical embodiment, they enable people to intuitively interact with and experience creative agents in physical reality. Vice versa, our reality allows for unbounded, complex robot behaviour to emerge in response to a rich stream of stimuli and their materialness. Each robot is equipped with speakers and two microphones, and maximises hedonic novelty through RL. By calculating a separate reward for each microphone, the robots steer towards more
interesting tunes. Similar to the **Digital Clockwork Muse** (Saunders & Gero, 2001d), they realise both the roles of author and critic through a simulation of ‘boredom’: as soon as their evaluation of other agent’s tunes falls below a manually defined threshold, they switch from listening to generating their own tunes. Saunders, Chee and Gemeinboeck (2013) later analyse the interaction dynamics between **Curious Whispers** and people. Over three observational studies, human participants were equipped with a synthesiser, allowing them to create tunes which the robots can perceive, evaluate and adopt, thus becoming part of the field contributing new artefacts into the creative domain. The participants understood that the intrinsically motivated robots cannot be commanded and became competitive in attracting them. They furthermore learned how the robots manipulate the tunes they gave them, thus building mental models of their robotic interaction partners.

Saunders (2011) extends the **Artificial Creativity** framework (Saunders & Gero, 2001b) to investigate how domain-specific languages emerge from the interaction of intrinsically motivated individuals within a creative system, and vice versa how language dynamics affect the system’s development. Individual agent behaviour is again driven by the maximisation of a *hedonic novelty* reward, with the explicit goal of grounding language in the subjective experience of an agent rather than through an extrinsic prior. As in the original account, agents create and share works with peers if it passes a reward threshold. But in the 2011 extension, agents also communicate an utterance as a descriptor along with their work. Saunders conducts several experiments in which agents generate and evaluate simple, coloured shapes of different sizes and communicate their work while adhering to the protocol of different *language games*. He finds that an increased preference for novelty, formalised through parametrisation of the hedonic function, yields a substantial growth in the variety of produced shapes, i.e. the *ontology*, but only a weak growth in the *vocabulary* used to describe the shapes, thus increasing *ambiguity* in the descriptions. He also reports that the more agents from two distinct domains with different vocabularies are moved together, the quicker a new domain is formed, but an unbalanced ratio causes disruptions. Finally, he finds quicker convergence to the same communication rate if one agent acts as a *teacher* who does not update their mappings between words and shapes.

Gemeinboeck and Saunders (2013) develop **Zwischenräume** (German: ‘in-between spaces’), a robotic art performance, to explore the potential synergies between robotic art and CC research. Amongst others, they highlight that CC can enrich art performances with open, non-determined modes of interactions. They use curiosity as an IM model specifically to this end, but also to increase the robot’s autonomy (cf. A.3) and goal-ownership (cf. A.2), and to promote shared or distributed agency within the creative act. **Zwischenräume** consists of a series of mobile robots mounted on rails inside the walls of a gallery. Each robot is equipped with a camera on a movable arm, a motorised hammer, and a microphone. The robots learn to control their position, camera arm and hammer based on maximising a set of distinct *novelty* rewards through **RL**, calculated separately from the microphone and camera input. To satisfy their need for novelty, the robots explore the wall, knock against it, and eventually punch holes into it through which they can then discover, study and respond to the human audience in the exhibition space. They thus realise
two distinct notions of creativity, in continuously introducing novel changes to the environment, and in contributing to the audience’s meaning-making. Saunders and Gemeinboeck (2014) develop Zwischenräume further into Accomplice, equipping the robots with more mobility, intricate chiselling patterns, and focussing more on the implicit communication between agents, e.g. through marks left in the wall. Both installations allow for distributed agency: ‘Rather than being invited to control the course of events, the audience is implicated in the material interventions of Accomplice, becoming accomplices in the works [sic] ongoing transformations’ (ibid., p. 3).

**Tackling the complexity of creative search (A.11)**

So far, we have mainly been looking at applications of IR to modelling and evaluating properties of creative products (Sec. 4.2.1). Next, we consider how such reward can be used in IM for modelling the creative process, e.g. in the form of exploratory and transformational creative search (Boden, 1990/2003; Wiggins, 2006a, 2006b). Vigorito and Barto (2008b) emphasise that creative search is often characterised by sparse extrinsic reward, e.g. ‘evaluative information may only be received upon completion of a creative endeavor’ (ibid., p. 1). This lack of guiding feedback is reflected in Wertheimer’s (1959) notion of productive thinking which requires ‘creating rather than recalling a solution’ (Peters, 1998, p. 836), and McGraw’s (1978) concept of a ‘heuristic solution’, where the steps to a solution algorithm must yet be discovered. If the creative process was only guided by such sparse extrinsic reward, it would be mostly blind (Vigorito & Barto, 2008a) and likely inefficient. In combination with the observation that ‘the size of this search space in most realistic domains is astronomical’, this ‘precludes any hope for success of blind trial-and-error processes in producing complex creative works’ (ibid., p. 135). A major reason to use models of IM in the previously discussed work is given by its ability to yield high performance on an extrinsically defined task in the presence of sparse extrinsic reward or its total absence (B.2). Furthermore, IM can give rise to exploratory behaviour (B.1) that is not only effective, but efficient. Researchers consequently leverage models of IM to tackle the complexity of creative search (A.11), albeit often only implicitly.

Vigorito and Barto (2008b) summarise that ‘successful creative search must (...) use self-generated, intermediate feedback as a surrogate for the sparse reward signals provided by the environment’ (ibid., p. 1, emphasis added). IR is used as such a surrogate when approximating human aesthetic judgement (A.5) in the search for artefacts that appeal to people (e.g. Saunders, 2009). But this specific case of implicit reward alignment is not commonly mentioned in the debate; instead, researchers focus on complementing rather than replacing extrinsic by intrinsic reward (IR), such that sparse extrinsic reward can be obtained in fewer steps. Vigorito and Barto (2008b) for instance propose using various competence-based models of IM (cf. Miorilli and Baldassarre, 2013; and Sec. 2.2.2) such as intrinsically motivated reinforcement learning (cf. Singh, Barto and Chentanez, 2005; and Sec. 2.2.4) to identify ‘behavioural building blocks that allow [an agent] to change the values of variables in its environment reliably by learning temporally and spatially abstract skills’ (Vigorito & Barto, 2008b, p. 1). This parallels Merrick, Saunders and Maher’s (2007) effectance
motivation and Colin et al.’s (2016) use of various models of curiosity to explore the space of options in hierarchical RL.

Learning complex action sequences to trigger changes in the search space represents only one means to support creative search, with another being efficient exploration: already Saunders and Gero (2001a) find that curiosity as a model of IM enables an agent to discover a larger variety of patterns than a random walk, thus allowing them to gain a better understanding of non-routine tasks. Colin et al. (2016) note that models of curiosity can tackle complex search spaces because they are not only better than random search, but also more efficient than an exhaustive search strategy. Finding a short-cut to such exhaustive search has previously been formulated by Wiggins (2006b) as a stepping stone towards modelling behaviour which is perceived as more creative (cf. Sec. 4.2.1). Exploration of a search space is even more efficient when multiple agents with different embodiments are at work. Inspired by Saunders and Gero’s (2001) Artificial Creativity framework, Linkola, Takala and Toivonen (2016) show, amongst others, that a society of curious agents with different memory sizes ‘can be more productive in generating novel artifacts than a single-agent or monolithic system’ (ibid., p. 8). In contrast to Saunders and Gero (2001b), they maximise a less powerful intrinsic novelty reward, but allow for agent self-criticism, veto power and voting ‘to collectively regulate which artifacts are selected to the domain’ (Linkola, Takala & Toivonen, 2016, p. 1).

Modelling Mini-Creativity (A.12)
The majority of related work aims at modelling big-c behaviour (cf. Kaufman and Beghetto, 2009; and Sec. 4.2.1), ignoring how such sophisticated creative behaviour arises from mini-c acts. Cohen (1989) relates to different types of adaptation to model the developmental continuum between such mini-c acts in children and more sophisticated forms of creative behaviour, including big-c acts, in adults. She argues that all but especially the first stage of creative behaviour require an individual to adapt to their environment by forming new mental connections that explain its workings. Crucially, Cohen deems ‘curiosity and pleasure in novelty’ (ibid., emphasis added) as well as play as particularly important for this p-creative act (Boden, 2003; and Sec. 4.1.1). A small body of related work has explicitly addressed the potential of IM to model mini-c creativity (A.12), drawing heavily on the concept of adaptation. Researchers exploit the capacity of intrinsically motivated behaviour to induce skill and model development (B.4) and to yield open-ended adaptation to different domains, agent embodiments and tasks (B.3), the latter resting on the domain and embodiment generality of IR (R.3).

Maher, Merrick and Macindoe (2005) leverage the potential of IM models to adapt an agent to new domains without domain-specific knowledge. Their focus is on designing an ‘intelligent room’ that adapts their behaviour in response to novel events happening inside. To facilitate such creative agency, they endow the room with an IM cascade, consisting of a model of curiosity that feeds into an effectance motivation. Both models operate on events as differences of sensor states, following actions performed by the room or its users. The hedonic novelty underlying curiosity is given by an earlier developed, event frequency cluster novelty detector (Kasmarik, Uther &
The room uses the previous action, event and reward to learn behavioural rules that are likely to trigger particularly interesting events, and the *effectance reward* reflects the confidence in each rule. Maximising effectance through RL enables the room to learn behaviours that induce interesting changes in the perceivable parts of its interior. Merrick, Saunders and Maher (2007) later extend this proposal by a *curious room* implemented as a society of curious agents (Saunders & Gero, 2001b) with different roles, some combining intrinsic and extrinsic reward. They hereby use curiosity explicitly to warrant open-ended adaptation to human behaviour: ‘Human activity is often characterised by creativity that leads to unpredictable changes in behavioural patterns. Consequently, it is difficult for system designers to predict in advance all the human behaviours that an intelligent environment may need to adapt to’ (Merrick, Saunders & Maher, 2007, p. 35).

Merrick (2008a) embraces a model of curiosity to enable a reconfigurable robot to adapt their behaviour to changes in their embodiment. Her goal is to teach designers creativity, reflection and imagination by observing the emergent behaviours from changing a robot’s sensors, actuators and body. She emphasises the need for a motivational function that is agnostic with respect to a change in sensors and actuators, and adopts an earlier developed curiosity model (Merrick & Maher, 2007) that maximises a *hedonic novelty* reward through RL. The reward is calculated for differences of sensor states, by first assessing their novelty with a hierarchical self-organising map, and then transforming it through a hedonic function. A study on Lego Mindstorms robot configurations yields behaviours which ‘encourage designers to play with different robot structures, reflect on the relationship between structure and behaviour and imagine new structures’ (Merrick, 2008a, p. 149).

Wiggins (2018) puts a curiosity-driven attention mechanism and the ongoing adaptation of memory toward higher compression at the basis of human everyday creativity and development. His cognitive architecture, *Information Dynamics of Thinking*, serves to explain spontaneous, value-less creativity, and is illustrated in the domains of music and speech. It models the human brain as an information-efficient, predictive processing (Clark, 2013) machine: ‘it predicts its world, so as to use information efficiently, and regularly re-represents it, so as to store information efficiently’ (Wiggins, 2018, p. 1). Both processes are driven by information-theoretic principles similar to Schmidhuber’s (2010) notions of *curiosity* and *compression*, respectively. External stimuli are only memorised after receiving conscious attention, conditional on being more *novel* than other stimuli. Memory in turn is regularly re-presented by chunking and linking stimuli sequences in a multi-dimensional, hierarchical statistical model, hereby exploiting new patterns in the data to minimise *model complexity*, i.e. maximise compression. Wiggins understands spontaneous creativity as *freewheeling predictions* in the absence of external stimuli, based on sampling from memory. In predicting a sequence that is novel relative to their memory, a system may e.g. generate new fragments of music. Wiggins predicts the developmental trajectory of a system with this architecture: Initially, it will imitate what is shown to them,

---

22 It is presently unclear how Wiggins’ (2018) architecture relates to hierarchical applications of the *Free Energy Principle* (Friston, 2010; and Sec. 2.2.4) to formalise predictive processing.
and exhibit *motor babbling*. Further on, the system will ‘learn the relation between motor signals and their proprioceptive or visual correspondents, and thus build a model of its own physical capabilities’ (Wiggins, 2018, p. 32).

### 4.2.2.3 Summary and Discussion

The preceding review has revealed a rich landscape of often co-dependent CC applications that, inspired by theories of (computational) creativity, leverage (emergent) properties of IR and intrinsically motivated behaviour. Here, we summarise our findings to answer our specific research questions.

The first question (RQ.3), ‘Why have IR and models of IM been used in CC?’, is answered in the form of the properties and the abstract applications that IR and IM models allow for, captured by our typology in Fig. 4.2. A specific IR function can be defined with or be attributed different semantics (R.1), and can thus be used to quantify novelty and value (C.1) as core components of creativity (A.1). IR is subjective and sensitive to an agent’s embodiment and situatedness (R.2) and thus allows for autonomous creative agency (A.3) to emerge from the perspectives of individual agents (A.2), potentially embedded in larger societies as a basis for modelling systems theories of creativity (A.9, A.10). By virtue of its potential alignment with extrinsic reward (R.4), IR can be used as a surrogate for sparse or unavailable extrinsic reward in creative systems (A.4). This includes human IR (C.2), which is at the basis of and thus allows for the approximation of some forms of human aesthetic judgement (A.5). Specific instances of IM models, i.e. a specific combination of intrinsic action-value and action-selection functions, can give rise to exploratory behaviour (B.1). Consequently, IM can be used as a heuristic traversal strategy to realise exploratory (A.6) and transformational creativity (A.8) in creative search. Since IR can align with extrinsic reward (R.4), intrinsically motivated behaviour can yield performance on extrinsically defined tasks in the absence of extrinsic rewards (B.2). It can therefore drive the creative process in complex search spaces with little or no extrinsic guidance (A.11), hereby tackling one of the biggest challenges of creative search. In combination with specific IR functions (cf. C.2), this can be exploited to autonomously create artefacts that appeal to people (A.7). The complexity of creative search is further reduced by mapping the search space and developing efficient traversal strategies, based on the potential of IM to induce skill and model development (B.4). This and the requirement of IM models to yield open-ended adaptation (B.3) have been leveraged to model different forms of mini-c acts (A.12).

The applications named above abstract from specific and often convoluted uses of IR and IM in related CC work, and partially answer our second specific research question (RQ.4): ‘How have IR and models of IM been used in CC?’ We complement this answer by analysing the characteristics of the underlying, specific studies with respect to existing schemata to classify CC research (cf. Sec. 4.2.1). Roughly one third of the identified studies investigate the benefits of IR and IM models through theory only, and two thirds apply them in simulation studies. The creative domains targeted are hereby well balanced, from creative behaviour and the development of language, through design and culinary creativity, to artistic, musical and scientific creativity. That
said, the majority of examples embraces a specific creative domain as an aid for the study of (human) creativity rather than as an engineering challenge. Further work bridges between this cognitive and the engineering perspective, and evaluates the use of CC systems in an arts context for both scientific and artistic benefit. Almost every proposal exploits IM for the modelling of autonomous systems, but more than half of the contributions also motivate the same model’s use for co-creativity\(^{23}\), with an even amount of work addressing co-creativity exclusively between machines or between machines and people. Almost half of the contributions reviewed here use IM to drive the behaviour of multiple agents. While most systems use an agent’s IR directly to drive their decisions, three systems propose an indirect usage, leveraging the reward or dynamics of intrinsically motivated agents to steer the behaviour of an overarching system that does not come with the embodiment required for intrinsically motivated agents, e.g. an evolutionary algorithm. Only a small majority of the existing studies assume a form of virtual embodiment, almost astride with work considering IM in physically embodied agents.

This typology crucially follows from a post-hoc analysis of existing work using our state-of-the-art definition of IM; most instances of existing work in fact appear unaware that the properties of the specific IM model they focus on also apply to a larger class of motivational models. When the notion of ‘intrinsic motivation’ is explicitly used, it is often considered synonymous with exploration (e.g. Grace & Maher, 2015). Indeed, almost all existing studies model some form of exploration-inducing curiosity, e.g. via (hedonic) novelty, learning progress or a surprise reward. While these can be summarised under the umbrella of knowledge-based models (Mirolli and Baldassarre, 2013; and Sec. 2.2.2), competence-based models such as effectance motivation and intrinsically motivated reinforcement learning, have rarely been investigated.

Crucially, not every model in the existing work meets all diagnostics of IM (cf. Sec. 2.2.3). We have evaluated their individual fit, but because many studies miss a full formalisation and provide ambiguous information with respect to the reward function used, our diagnoses in Tbl. 4.1 should be taken with a grain of salt. Since a detailed discussion of all models is out of scope, we support this information with only two examples. The surprise reward proposed by Macedo and Cardoso (2001a) is agent-centric, but does not fit any other diagnostic of IR. This is mostly due to its symbolic nature: sensing is reduced to directly perceiving or imagining products, which are memorised as graphs of symbolic knowledge. The approach thus relies on a closed knowledge base, which rules out freedom of semantics and embodiment sensitivity, and severely limits open-endedness. The hedonic learning progress reward proposed by Smith and Garnett (2012b) in contrast fulfils all diagnostics of IM but the behaviour resulting from maximising it would not be open-ended. This is because it measures the change of model weights rather than the difference of prediction error and a maximising agent can thus get stuck with ‘noisy TVs’ (Burda, Edwards, Storkey et al., 2019; and Sec. 2.2.4). Due to these individual shortcomings, not all models in related work can leverage all reasons to use IM in CC.

\(^{23}\) Some only talk of co-creativity when multiple systems contribute to the same product; we deem it sufficient when they mutually influence their potentially distinct creative processes.
The fit of the reviewed models to our working definition of IM only conveys little information about their relationship to the new empowerment-based models developed in Ch. 6 and 7 of this thesis. We next highlight the strength of our contributions by relating them to the reviewed work.

4.2.3 Contextualising Our Contributions

We finally outline how our applied contributions in Ch. 6 and 7 relate to and advance the existing work identified in our systematic review. We extend these references beyond work that uses IR and IM in the individual chapters.

Our systematic review has revealed a low diversity in the IRs and IM models used in existing work and a focus on few models that yield exploratory behaviour. This is not surprising, given the origins of the IM concept in the observation of exploratory behaviour (cf. Sec. 2.1.1), the early introduction and continuing popularity of curiosity-based models of IM in AI more generally (cf. Sec. 2.2.4), the dominance of novelty and surprise in the definition of creativity (cf. Sec. 4.1.1), the central role of exploration in the mechanisms of exploratory and transformational creativity (cf. Sec. 4.2.2), and creativity studies’ emphasis on exploration in discussions of the IM-creativity relationship (cf. Sec. 4.1.2). Our typology highlights how models of IM more generally, rather than e.g. curiosity as a specific model, can benefit CC, thus paving the way for leveraging other models of IM. We demonstrate this opportunity by applying empowerment and empowerment maximisation (EM, cf. Ch. 3) as a different IR and IM model across the following two contributions.

Co-creativity is a particularly interesting application domain for models of IM due to the challenging complexity and open-endedness introduced by interaction partners, and people in particular. More than half of the reviewed work focusses on driving co-creative behaviour, letting interaction emerge from the IM of the involved agents (e.g. Macedo & Cardoso, 2001b; Saunders & Gero, 2001b; Saunders et al., 2010). Crucially, the interaction partners are never formalised as agents in the calculation of IR, but always treated as mere parts of a sophisticated environment. The outcome is a vast, open space of possible interaction dynamics, only influenced by the embodiment and situatedness of the individual agents. While this openness may be desirable for the study of (human) creativity from a cognitive perspective, it poses a tricky challenge from the engineering perspective: how can we leverage the benefits of IM in co-creativity, while limiting the emerging interaction dynamics to those that are desirable for the interaction partners, either human or machine? In Ch. 6, we introduce a model of social intrinsic motivation to tackle this challenge. By calculating and combining different variants of empowerment as IRs, based on explicitly modelling the behaviour of other agents within a shared environment, our formalism allows us to control the emergent social interaction dynamics: we can interpolate between the extremes of supportive to antagonistic behaviour, while still leveraging the benefits of IM. We motivate our model in CC through thought-experiments, and motivate as well as evaluate it via simulation experiments in the domain of videogame AI.
While the previous contribution focusses on driving the process of mini-c acts, our contribution in Ch. 7 concentrates on the generation and evaluation of sophisticated, interactive products, i.e. artefacts that can be actively changed through intervention by the audience such as a kinetic sculpture or a happening. Evaluating such dynamic artefacts is particularly complicated if the interaction outcome can mostly be captured in terms of subjective experiences. IR functions are promising candidates for evaluating such experiences, as they are inherently subjective. Furthermore, they are independent of specific domain knowledge and embodiments, and can thus be used across different domains to model different audiences. Saunders and Gero (2004) were the first and only to use IR in the evaluation of interactive experiences. Our contribution takes their indirect approach to using IR further by not only evaluating interactive artefacts through IR to predict how people would experience them, but also using the outcome to generate new interactive artefacts. In contrast to the existing approach, we leverage the IR of artificial agents following a policy that may differ from the one that optimises the reward. This crucially does not violate the diagnostic of agent-centricity (Sec. 2.2.3) as long as the intrinsic reward is calculated only on the internal components of a model describing the embodiment of the experiencing agent. In summary, we propose, instantiate, and evaluate a blueprint for autonomously generating human-appreciable interactive artefacts based on an off-policy, IR estimation of human experience. We again motivate our proposal through CC thought-experiments, and demonstrate it in the domain of videogame AI.

In this chapter, we have argued that models of IM can advance core areas of CC research. Based on a systematic review of related work, evaluated against a strong theoretical basis, we devised a typology that answers why we should embrace models of IM in CC (RQ.3), and how such models have been leveraged so far (RQ.4). We next use this typology to guide a systematic review of how models of IM can advance computational game creativity (Liapis, Yannakakis & Togelius, 2014) as a CC sub-field and part of videogame AI. Based on our insights, we motivate how our contributions to videogame AI in Ch. 6 and 7 can benefit game players, designers and AI engineers.
Videogame AI is a striving area of research, with a strong demand for innovation by games industry and players. In this chapter, we seek to support our second overarching research question, ‘Can IR and models of IM advance videogame AI?’ (RQ.2), by answering two specific research questions with respect to existing work:

RQ.5 Why have IR and models of IM been used in videogame AI?

RQ.6 How have IR and models of IM been used in videogame AI?

In the spirit of computational game creativity (Liapis, Yannakakis & Togelius, 2014), we raise a third question to understand whether some if not all instances of related game AI work could also be considered examples of CC:

RQ.7 How do existing applications of IR and IM models in videogame AI and CC overlap?

An affirmation would allow us to study CC ‘within and for computer games’ (ibid., p. 47) and thus indirectly support our first overarching research question: ‘Can IR and models of IM advance CC?’ (RQ.4).

Game AI as we conceive it should ultimately benefit game engineers, designers and players. To understand how models of IM could advance this goal, we take the perspective of game design and games user research in Sec. 5.1 to understand what constitutes games and (game)-play, and what makes games intrinsically motivating for people. We then leverage these insights in Sec. 5.2 to introduce and motivate existing game AI work using models of IM and answer RQ.5 as well as RQ.6. We constrain this non-exhaustive but systematic review with the previous definition of games and our working definition of IM models from Sec. 2.2.3. We moreover inform our review by existing taxonomies of game AI research. We end up with a typology of applications of IM in game AI, which we compare with the CC typology from Ch. 4 to answer RQ.7. We eventually motivate and contextualise our applied contributions by relating to the existing work.

Our main contribution in this chapter is an elaborate, systematic review of existing work employing models of IM in game AI. Crucially, related surveys covering IM models only consider videogames as one of many benchmarks for artificial general intelligence. We in contrast also highlight the role of videogames as cultural artefacts and examine how IM models have advanced game AI to benefit game engineers, designers and players. We have already contributed to the state-of-the-art through a review of IM models in simulation-based game testing (Roohi et al., 2018). This chapter extends this review by considering all areas of game AI (cf. Yannakakis & Togelius, 2018, pp. 259-260). We furthermore provide a stronger theoretical grounding by employing our working definition of IM models (Sec. 2.2.3). While this constrains the body of related work further, it allows us to focus more on the
unique benefits that this class of models offers for game AI. By combining both a computational and non-computational perspective, we contribute a unified view of IM across game design, games user research and game AI that may foster future interactions between these disciplines.

5.1 INTRINSIC MOTIVATION IN VIDEOGAMES

We can consider at least two roles of IM in the context of videogames: a person can be intrinsically motivated to play, or to design a game. Our primary focus in the next two sections is on the motivation of players. But since play is the basis for assessing a particular game design, the way designers use the IM concept tells us a lot about its function in play. We consequently draw on both game design (Sec. 5.1.1) and games user research (Sec. 5.1.2) to better understand the role of people’s IM when playing a game, and to reveal why and when models of IM can be exploited for various kinds of game AI. But what do we mean by ‘play’, by ‘videogames’ specifically and by ‘games’ more generally?

‘Play’ is arguably the most ambiguous (cf. Sutton-Smith, 2009) of these concepts, and it is closely related to IM. The philosopher Santayana writes:

‘Play is whatever is done spontaneously and for its own sake.’
(Santayana, 1896, p. 19)

The game designer Schell (2019) criticises this definition because play can often be planned, and spontaneity is thus not a strict requirement. But he also praises it by highlighting the doing of something for its own sake as a persistent characteristic of play. This clearly resonates with the definition of IM as the ‘doing of an activity for its inherent satisfactions’ (Ryan & Deci, 2000a, p. 56). Vice-versa, intrinsically motivated behaviour is often described as ‘playful’ (e.g. Amabile, 1979; Ryan & Deci, 2000a). Despite these similarities, the literature is ambiguous with respect to whether play is an intrinsically motivated activity: By stating ‘If we don’t like to do it, it probably isn’t play’ (Schell, 2019, p. 28), Schell reduces play to an activity that is merely volitional, and could thus also be extrinsically motivated through an internal locus of causality. We complement this definition by Salen and Zimmerman’s account:

‘Play is free movement within a more rigid structure.’
(Salen & Zimmerman, 2004, p. 304)

This emphasises that play is contingent on constraints within a reference system: ‘Play is an expression of the system, one that takes advantage of the space of possibility created from the system’s structure’ (Salen & Zimmerman, 2004, p. 304). This system could be natural, or it could be artificially constructed in the form of a ‘game’.

Similar to the definition of ‘creativity’ (Sec. 4.1.1), what constitutes a game is subject to a long and ongoing debate, with some considering games as ‘forever indefinable or ungraspable’ (cf. Juul, 2003, p. 43). Wittgenstein even motivates his concept of family resemblance (cf. Sec. 4.1.1) through games (ibid., §66-67). The struggle to find a definition has been summarised by Stenros (2017), who examines 63 definitions of games over the past seven decades to identify what they agree and disagree on. Since the exact differences do not
matter as much for this thesis, we adopt a definition by Juul (2003) that is informed by a similar but smaller analysis of existing accounts:

‘A game is a rule-based formal system with a variable and quantifiable outcome, where different outcomes are assigned different values, the player exerts effort in order to influence the outcome, the player feels attached to the outcome, and the consequences of the activity are optional and negotiable’ (Juul, 2003, p. 35).

We constrain our scope further by adopting Juul’s definition of ‘videogames’ as ‘all games played using computer processing power’ (ibid., p. 30). We finally contrast games with toys. According to Schell, ‘Games have rules. Toys do not have rules. Rules are definitely one of the defining aspects of games’ (ibid., p. 31). ‘Toys are fun to play with for their own sake. In contrast, games have goals (...’ (ibid., p. 119). We consider toys as systems of constraints that afford play as free movement within, but that lack goals to guide this movement. Games are played, while toys are played with (ibid., p. 27).

Schell’s remark that toys are played with ‘for their own sake’ again resonates with the definition of IM (Ryan & Deci, 2000a). But is such play really an intrinsically motivated activity? How is playing a game, where goals are present, different from playing with a game, i.e. using it as a toy, for its own sake? We next investigate the role of IM in play through the lens of game design, focussing on the general IM concept rather than specific theories. By resolving ambiguity in the literature, we address the above questions and argue why models of IM can benefit videogame AI.

5.1 Intrinsic Motivation in Game Design

To diagnose intrinsically motivated behaviour, psychologists try to rule out that a person acts for a separate outcome. Game-playing represents a particularly useful activity to understand when people are intrinsically motivated. This is because gameplay, in contrast to many other activities (cf. the examples in Sec. 2.1.1), is mostly closed with respect to separate outcomes, and games can be simplified to a high degree without losing expressivity.

Games are separated from larger reality through what Huizinga (1950) calls the magic circle, i.e. a ‘boundary – or frame – that defines the game in time and space’ (ibid., p. 10). When a person starts playing a game, they cross over this boundary to adopt the artificial behaviours and rituals of a game. During the game, the magic circle persists until the game concludes (Salen & Zimmerman, 2004, pp. 332-333). In their influential book on game design, Salen and Zimmerman stress that the magic circle establishes rules and goals apart from ‘ordinary life’. They furthermore follow from the self-contained nature of the magic circle that games are autotelic activities, i.e. they have an end or purpose in themselves, because they ‘contain their own meanings and provide their own goals’ (ibid., pp. 332-333). While this is a defining characteristic of games, it does not imply that a game cannot also have separate consequences: people engage with games to win prizes in e-sports competitions or to impress their friends at the arcade. But, as part of his game definition, Juul (2003) requires these consequences to be both optional and negotiable: ‘I suggest that games are characterized by being activities with
**5.1 INTRINSIC MOTIVATION IN VIDEOGAMES**

**negotiable** consequences: A specific playing of a game may have assigned consequences, but a game is a game because the consequences are *optionally* assignable on a per-play basis. That games carry a degree of separation from the rest of the world follows from their consequences being *negotiable* (Juul, 2003, p. 34, emphasis added). Salen and Zimmerman note that games as autotelic activities are done ‘not with the expectation of some future benefit, but simply because the doing itself is the reward’ (ibid., p. 332). Similarly, Caillois (1961) describes a game as ‘[…] an activity which is essentially: Free (voluntary), separate [in time and space], uncertain, unproductive, governed by rules [and] make-believe’ (ibid., pp. 10-11). This resonates strongly with Ryan and Deci’s (2000) definition of IM; and since games must be autotelic by definition, we may consequently be tempted to conclude that any game is also intrinsically motivating.

Unfortunately, the game design literature is too ambiguous to confirm or reject this idea. Salen and Zimmerman seem to support it by noting that ‘games are, to a greater or lesser extent, pursued for their own sake, for their own intrinsic stimulation’ (ibid., p. 333) and that, ‘although there are always some extrinsic reasons for [game] play, there are always intrinsic motivations as well’ (ibid., p. 333, emphasis added). This stands in slight conflict with Juul’s remark1 that games can ‘be played with or without real-life consequences’ (Juul, 2003, p. 35, emphasis added). We also note that these references relate to ‘stimulations’ or ‘pleasures’ as inherent properties of games, rather than relational properties between a game and a specific player: ‘all games also provide intrinsic, autotelic pleasures that are significant only within the artificial meanings that the game creates’ (Salen & Zimmerman, 2004, p. 360). We thus find that in game design, the terms ‘autotelic’ and ‘intrinsically motivating’ are often used synonymously to describe games, the player as the subject of motivation is left out, and there is no consensus on the relationship of IM and play.

We resolve this ambiguity in the literature with the following thought-experiment: Consider playing a game in which you find yourself in a small cabin situated within a vast and diverse world that is discernible through the windows. The goal is to score as highly as possible, and the game increases your score the longer you stay in the cabin. The described experience is autotelic, in that it establishes a goal separate from, and with no bearing on, reality. Now ask yourself: would you keep playing this game without being told to do so? Most people would probably say no; they would deem the game unenjoyable, and not play it in the absence of external pressure. This imaginary ‘free choice’ experiment (cf. Sec. 2.1.2) highlights that most people would not be intrinsically motivated to play this game. Thus, **being autotelic is not sufficient for a game to be intrinsically motivating**.

In order to understand what is missing for a game to be intrinsically motivating, we have to consider the relationship of the game and a specific person, from the perspective of that person rather than the game. The person as an active agent will only play the game, i.e. they will only become motivated to work toward and achieve game-internal but player-external goals, if this

---

1 Juul (2003) notes that the requirement for games to have optional, separate consequences is an *ideal*, as often such consequences are beyond our (conscious) control. I.e. a ‘pure autotelic’ game as in our thought-experiment probably does not exist outside lab conditions.
activity is intrinsically rewarding. Consider altering the reward mechanic in our thought-experiment game so that the score is increased when the player successfully completes quests that require leaving the cabin and exploring the game world. We believe that many more people would play this version of the game without external pressure, and could thus be considered intrinsically motivated. This is because engaging in these quests would allow them to satisfy their need to experience unfamiliar situations, i.e. their curiosity. For a person to be intrinsically motivated to play a game, progress on autotelic goals must align with player-intrinsic reward.

Even if a player engages with a game without external pressure, we cannot assume an alignment between the game’s goals and the player’s IRs. This is because a person might not be intrinsically motivated to play the game, but to play with the game, i.e. using it as a toy. This is possible because ‘many games are built on top of toys’ (Schell, 2019, p. 119). Schell proposes to guide game design by asking: ‘When people see my game, do they want to start interacting with it, even before they know what to do? If not, how can I change that?’ (ibid., p. 90). Minecraft (Mojang & Microsoft Studios, 2009) for instance has long been a toy, and was only turned into a game later by adding a survival mode. When engaging with a toy, a person becomes the designer of their own experience, leveraging the free movement the toy affords for the satisfaction of their IMs (cf. Bogost, 2016). Costikyan (2002) notes that the game SimCity (Maxis, 1989) ‘works because it allows players to choose their own goal, and supports a wide variety of possible goals’ (Costikyan, 2002, p. 13). IM explains why people e.g. build vast structures in Minecraft, even if this does not necessarily contribute to their survival and thus to the game’s goal. We summarise that a person can be intrinsically motivated to play with a game, rather than playing the game, by ignoring the game’s goals.

There is a reason why the game design literature is ambiguous with respect to the role of IM: games are designed for no other reason but to be played: ‘crossing into the magic circle as well as maintaining its existence, represent two of the chief challenges of designing meaningful play’ (Salen & Zimmerman, 2004, p. 333). A game must by definition be autotelic (ibid., pp. 332-333) and separate outcomes must be optional (Juul, 2003). In its purest form, a game is thus only played when people become intrinsically motivated to do so. Since game design strives for continuous player engagement, games are implicitly designed to be intrinsically motivating for as many people as possible. In other words, game design implicitly entails aligning game-internal and yet player-external goals with player-intrinsic reward. Traditionally, IM is often associated with curiosity (cf. Sec. 2.1.2), and the effect of curiosity on increasing player’s engagement is so pronounced that Schell even dedicated a game design lens to it: ‘To use this lens, think about the player’s true motivations – not just the goals your game has set forth, but the reason the player wants to achieve those goals’ (ibid., p. 30, emphasis added). We crucially do not claim that IM is the only mechanism motivating human game-playing, but it is likely one of the key factors in fostering player engagement.

Our thought-experiment reveals the central role of IM in the design and play of and with games. But these insights only rest on the general IM concept, and do not yet reveal how much IM really contributes to human play, and what specific models are at work. In the next section, we complement the
present theoretical argument by consulting empirical studies from games user research that demonstrate the role of specific psychological IMs in driving human play. This allows us to understand why certain IM models might be better suited for specific game AI applications than others, and to motivate the use of empowerment as IR in our novel models and applications.

5.1.2 Intrinsic Motivation in Games User Research

Games user research employs empirical methods from various disciplines such as human-computer interaction and psychology (Drachen, Mirza-Babaei & Nacke, 2018, p. 1) to help game designers and engineers creating ‘better gameplay experiences by finding weaknesses in the design and structure of games’ (ibid., p. 2). We have previously highlighted that fostering continuous player engagement is a primary concern in game design, and pointed out IM as facilitator of such engagement. It is thus not surprising that studies on the relationship of IM and human play feature prominently in games user research. The focus of these studies is on specific theories of IM rather than the general concept, and we believe that they can highlight opportunities for exploiting specific models of IM in game AI. We delineate the scope of research, but do not provide an exhaustive review.

Malone proposes as early as 1981 to foster IM in videogames by designing for challenge and curiosity, amongst other factors. He understands challenge as the driving force in the development of competence and feelings of efficacy as proposed by the theory of effectance motivation (White, 1959; Harter, 1978). He furthermore draws on flow theory (Csikszentmihalyi, 1990) to identify features that characterise challenging activities. Finally, he adopts Berlyne’s (1960) theory of curiosity that is based on Hunt’s (1965) optimal incongruity theory of IM. Malone proposes to modulate challenge as the outcome uncertainty of game goals, and to increase player’s curiosity by making them believe that their knowledge structures are incomplete or inconsistent. But contrary to most studies in games user research, he does not measure these factors quantitatively (cf. Denisova, Guckelsberger & Zendle, 2017; Denisova et al., 2020). Yannakakis and Hallam (2004) complement Malone’s work by turning his qualitative factors into objective2, quantitative metrics of challenge, curiosity and fantasy in predator-and-prey games, and combine them into a real-time measure of a game’s entertainment value. In a subsequent user study, Yannakakis and Hallam (2007) show that their measure is highly correlated \((r = 0.44, p < 0.001)\) with the judgement of human players on the entertainment value of different variants of a Pac-Man (Namco, 1980) clone. This is within our scope because games user researchers have used fun and enjoyment to operationalise the strength of IM.

Klimmt and Hartmann (2006) as well as Klimmt, Hartmann and Frey (2007) investigate more closely the effect of effectance on player’s perceived enjoyment in videogames. They understand effectance as ‘receiving immediate, direct feedback on one’s action and of influencing the game world’ (ibid.,

---

2 These metrics do not qualify as intrinsic reward according to our working definition in Sec. 2.2.3, because they are not computed from the subjective perspective of an agent and thus violate the basic diagnostic of agent-centricity, amongst others.
Based on player’s ratings of effectance and enjoyment on a dedicated questionnaire, their findings confirm effectance as one of potentially many factors impacting game enjoyment and thus IM: ‘players enjoy watching the results of the actions they perform, and their fun declines if these efficacy experiences are limited’ (Klimmt, Hartmann & Frey, 2007, p. 847). They investigate situations in which the game controls only irregularly affect the game, but they do not cover the case where players’ effectance is explicitly constrained by a game’s mechanics, e.g. while a character is in the air after jumping. They also investigate the effect of control as the player’s capacity to anticipate the game’s dynamics and to influence them according to their goals, but do not find a significant interaction effect with enjoyment. They argue that the relationship of control and enjoyment is more complex, because challenges that contribute to enjoyment often constrain control.

While the previous studies rely on a variety of IM theories (cf. Sec. 2.1.2), games user research has arguably been dominated by self-determination theory (Ryan & Deci, 2000b). Ryan, Rigby and Przybylski (2006) have applied this theory to videogames in order to model player motivation across specific games, based on the assumption that ‘players of all types seek to satisfy psychological needs in the context of play’ (ibid., p. 349). Their hypothesis is that ‘games are primarily motivating to the extent that players experience autonomy, competence and relatedness while playing’ and that satisfaction of these needs ‘should thus predict subsequent motivation to play’ (ibid., p. 348). To probe their hypotheses, they have developed the Player Experience of Need Satisfaction scale by translating items from existing self-determination theory questionnaires to the games context. Applying this scale, they find that perceived in-game autonomy and competence are associated with game enjoyment and allow to predict future engagement. They recommend designers to afford more player autonomy by allowing for ‘flexibility over movement and strategies, choice over tasks and goals’ and give ‘feedback rather than to control the player’s behavior’ (ibid., p. 349). Furthermore, they suggest to increase competence by offering intuitive game controls that can be readily mastered, and ‘tasks within the game [that] provide ongoing optimal challenges and opportunities for positive feedback’ (ibid., p. 349).

The same questionnaire and self-determination theory as a formal framework has been extensively used in further studies. Przybylski, Rigby and Ryan (2010) for instance show empirically that the need satisfaction of competence, autonomy and relatedness can contribute to IM and serve as a robust predictor across player demographics, game genres and content. Peng et al. (2012) add further support with experiments manipulating game features of an exergame with respect to each of these three basic needs. They find that such manipulation allows to predict need satisfaction, game enjoyment and motivation for future play, amongst others, and that the needs for autonomy and competence serve as mediators between game features and ongoing engagement. Abuhamdeh, Csikszentmihalyi and Jalal (2015) explicitly relate self-determination (Ryan & Deci, 2000b) and flow theory (Csikszentmihalyi, 1990) in competitive videogames. The concept of optimal challenge is central to flow theory, and cognitive evaluation theory as a part of self-determination theory proposes that optimal challenges are enjoyable because they maximise perceived competence. Abuhamdeh, Csikszentmihalyi and Jalal (2015) adopt
Malone’s (1981) early idea of linking challenge to outcome uncertainty of goals, and investigate how outcome uncertainty and perceived competence affect players’ IM. Their participants preferred games with high outcome uncertainty, mediated by the feeling of suspense, over those that merely maximised perceived competence. This suggest that optimal challenges, as suggested by flow theory, yield IM because they promote outcome uncertainty experienced as suspense, and that maximising competence and autonomy alone, as proposed by cognitive evaluation theory, might not be enough to predict IM and engagement.

The previous findings are complemented by research into player typologies. Hamari and Tuunanen (2014) have proposed a meta-synthesis of existing typologies into the five dimensions of (i) achievement, (ii) exploration, (iii) sociability, (iv) immersion and (v) domination. We can associate these player types with different theories of IM: (i) theories of effectance motivation, personal causation and the concept of competence in self-determination theory, (ii) theories of curiosity, (iii) the notion of relatedness in self-determination theory, (iv) flow theory and (v) self-determination theory’s need for autonomy in terms of not being dominated or controlled by others. This allows us to make several points. Firstly, we can retrieve almost every established psychological theory of IM in games user research. This not only highlights the wide scope of this field, but also the diversity of IM in player motivation. Player typologies cover both intrinsic and extrinsic motivations in play, but we believe that this association still shows the importance of IM in human play. We secondly find that not all player types can be explained in terms of a single IM theory. This supports that players are motivated by a complex interplay of intrinsic needs, and challenges whether e.g. self-determination theory (Ryan & Deci, 2000) alone is sufficient to explain player motivation. We finally note that this association highlights inter-subjective differences in player motivation. As a consequence of this heterogeneity, Bartle stressed that ‘successful games must provide gratifications for all (...) player types’ (Bartle, 2004; as paraphrased by Ryan, Rigby and Przybylski, 2006, p. 348).

We find that games user researchers have considered a broad spectrum of psychological theories of IM (cf. Sec. 2.1.2), but, as in the game design literature (cf. Sec. 5.1.1), we also notice ambiguity in the use of the IM concept. Ryan, Rigby and Przybylski (ibid.) for instance put forward self-determination theory hypotheses that can be applied ‘both at the level of the player making choices between gaming products, and the motivation of a player while “in character” within a particular gaming context’ (ibid., p. 349). They thus mix up a person’s IM in playing a game, and in other activities outside the magic circle (Huizinga, 1956). The discussion of IM in games is further complicated because definitions of core concepts such as effectance, competence, autonomy and others are very vague and change frequently across publications. Finally, some concepts are very high-level and overlap between theories, e.g. competence (Ryan & Deci, 2000b) with effectance (Klimmt & Hartmann, 2006; Klimmt, Hartmann & Frey, 2007) and optimal challenge (Abuhamdeh, Csikszentmihalyi & Jalal, 2015). Tyack and Mekler (2020) summarise the state-of-the-art of self-determination theory research in games, and point out open questions.

Considered coarsely, we yet deem these findings very valuable; they not only empirically support the importance of IM in human game-playing, thus
complementing our theoretical argument in Sec. 5.1.1, but also highlight specific candidate theories of IM as particularly relevant. Also, the wealth of results highlights the challenge that human game-playing can likely not be explained through a single theory of IM. Based on our mapping from player types to motivations, and the need for successful games to attract different player types (cf. Bartle, 2004), we also conclude that commercial games, unless targeting a player niche, must provide multiple ‘hooks’ for different IMs to attach to. In the next section, we use these insights to motivate existing applications of IM models across different areas of videogame AI, and to contextualise our own contributions to this field.

5.2 REVIEW OF INTRINSIC MOTIVATION IN VIDEOGAME AI

Equipped with the key insights on the role of IM in game design and human play laid out in the previous section, we are almost ready to approach this chapter’s research questions. As final requirement, we provide a more principled account of videogame AI in terms of its goals and stakeholders.

Yannakakis and Togelius define videogame AI as ‘the study of AI in and for games’ (ibid., p. 4, emphasis added). We adopt their definition, because it uncovers widespread ambiguity in the use of the term. On the one hand, new AI techniques are evaluated in games as benchmarks for artificial general intelligence. Games here serve as a means to an end, with advances benefiting a wide range of stakeholders and applications, but typically not games specifically. On the other hand, AI is developed for games, i.e. to ultimately benefit game engineers, designers and players (cf. ibid., pp. 262-264). We distinguish four core areas of the latter kind of game AI research by modifying a taxonomy originally put forward by Yannakakis and Togelius (ibid., pp. 259-260): (i) the design of (game-)playing agents, (ii) the engineering of non-player character (NPC) behaviour, (iii) player experience and behaviour modelling, and (iv) procedural content generation (PCG).

In Sec. 5.2.1, we use this account of game AI to answer our three research questions via a systematic review of existing work exploiting models of IM for game AI. In Sec. 5.2.2, we eventually draw on our findings to contextualise and motivate our novel IM models and applications in Ch. 6 and 7.

5.2.1 Systematic Review

We answer our first two research questions, ‘Why have IR and models of IM been used in videogame AI?’ (RQ.5) and ‘How have IR and models of IM been used in videogame AI’ (RQ.6), through a systematic review of research employing models of IM in videogame AI. We inform and structure this review based on the preceding account of game AI, and by our working definition of IM models (Sec. 2.2.3). In contrast to other reviews, we consider both meanings of game AI, and all areas of research on the use of AI for games. Based on our findings, we derive a typology of abstract applications of a taxonomic account of videogame AI as ‘the study of AI in and for games’ (ibid., p. 4, emphasis added). We adopt their definition, because it uncovers widespread ambiguity in the use of the term. On the one hand, new AI techniques are evaluated in games as benchmarks for artificial general intelligence. Games here serve as a means to an end, with advances benefiting a wide range of stakeholders and applications, but typically not games specifically. On the other hand, AI is developed for games, i.e. to ultimately benefit game engineers, designers and players (cf. ibid., pp. 262-264). We distinguish four core areas of the latter kind of game AI research by modifying a taxonomy originally put forward by Yannakakis and Togelius (ibid., pp. 259-260): (i) the design of (game-)playing agents, (ii) the engineering of non-player character (NPC) behaviour, (iii) player experience and behaviour modelling, and (iv) procedural content generation (PCG).

In Sec. 5.2.1, we use this account of game AI to answer our three research questions via a systematic review of existing work exploiting models of IM for game AI. In Sec. 5.2.2, we eventually draw on our findings to contextualise and motivate our novel IM models and applications in Ch. 6 and 7.

5.2.1 Systematic Review

We answer our first two research questions, ‘Why have IR and models of IM been used in videogame AI?’ (RQ.5) and ‘How have IR and models of IM been used in videogame AI’ (RQ.6), through a systematic review of research employing models of IM in videogame AI. We inform and structure this review based on the preceding account of game AI, and by our working definition of IM models (Sec. 2.2.3). In contrast to other reviews, we consider both meanings of game AI, and all areas of research on the use of AI for games. Based on our findings, we derive a typology of abstract applications of

---

3 In contrast to Yannakakis and Togelius (2018, pp. 259-260), we consider the engineering of player- and non-player behaviour separate categories. We furthermore write (game-)playing to stress that this class subsumes research on playing games and on playing with games.
IM models in videogame AI, and of (emergent) properties of IR and intrinsically motivated behaviour that these applications leverage. We finally answer RQ.7, i.e. ‘How do existing applications of IR and IM models in videogame AI and CC overlap?’ by comparing the first typology with its CC analogue derived in Ch. 4. This allows us to decide whether the abstract applications identified in related game AI work can be considered instances of computational game creativity (Liapis, Yannakakis & Togelius, 2014).

5.2.1.1 Method

We filter related work on models of IM in game AI through a three-step process, the first two steps being similar to the strategy employed for our review of related CC research (Sec. 4.2.2). We (i) identify and coarsely filter candidate literature by their title and abstract, then (ii) extract qualifying work based on in-depth reviews, and finally (iii) only select those publications that are of direct relevance for motivating our work in Ch. 6 and 7.

For our first step (i), we manually4 combed through relevant conference proceedings and journals publishing work on each of the two views on game AI. For work employing AI for games, we consulted the Foundations of Digital Games Conference (FDG) (2009-2019); the Artificial Intelligence for Interactive Digital Entertainment Conference (AIIDE) (2005-2019) and the Conference on Computational Intelligence and Games (CIG) (2005-2018), which has recently been renamed to the Conference on Games (CoG) (2019). We also considered the Experimental AI in Games (EXAG) workshop as part of AIIDE (2013-2019); the Computational Creativity and Games Workshop (CCGW) as part of the International Conference on Computational Creativity (ICCC) (2015-2017); the Symposium in AI & Games as part of the convention of the Society for the Study of Artificial Intelligence and Simulation of Behaviour (AISB) (2009-2019); and the Computer Games Workshop (CGW) at the International Joint Conference on Artificial Intelligence (IJCAI) (2013-2018). We furthermore took into account journal articles in Computers in Entertainment (CiE) (2003–2018); Games (MDPI, 2010-2019) and in Transactions on Computational Intelligence and AI in Games (T-CIAIG) (2009-2017), which has recently been renamed to Transactions on Games (ToG) (2018, 2019). Specifically for research employing AI in games as benchmark, we also included the Conference on Neural Information Processing Systems (NeurIPS) (2001-2019), the International Conference on Machine Learning (ICML) (1996, 2002, 2004-2019), and the International Conference on Learning Representations (ICLR) (2013-2019) into our search. We identified further related work by following up references in the initial list of candidate publications.

For our second step (ii), we have filtered the list of candidates based on in-depth comparisons against both, Yannakakis and Togelius’ definition of videogame AI (2018, p. 4), complemented by Juul’s (2003) definition of videogames (Sec. 5.1), and our working definition of IM models (Sec. 2.2.3). We crucially do not consider any work that exclusively focusses on game theory and consequently exclude e.g. the work of Jaques et al. (2019). We leave out

4 In previous joint research (Roohi et al., 2018), we have identified related work via web search for the keywords ‘intrinsic motivation’ and ‘videogames’, amongst others. Due to the ambiguous usage of these concepts, this approach returns many false positives while missing out on relevant papers that do not use these keywords. We thus use a different approach here.
### Table 5.1: Reviewed work using IR and IM in videogame AI. The ○ and ● circles represent (un-)fulfilled criteria, and ◊ indicates a topic that has only briefly been touched on. We use ✔ where there is not enough or conflicting information on a criterion. ¹The applied part of this contribution violates our definition of IM models and we consequently only consider the theoretical motivation. ²We only relate to the sub-section on ‘Intrinsically Motivated RL for Content Generation’ here, as models discussed in other sub-sections violate our definition of IM.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Year</th>
<th>Identifier</th>
<th>Game AI Type</th>
<th>Reward Properties</th>
</tr>
</thead>
<tbody>
<tr>
<td>Merrick and Maher</td>
<td>2006</td>
<td>Motivated NPCs in Persistent Videogame Worlds</td>
<td>A D ○ ● ○ ○</td>
<td>Hedonic Novelty Max. (RL) ● ○ ○ ○</td>
</tr>
<tr>
<td>Merrick and Maher</td>
<td>2007</td>
<td>Adaptive NPCs in Open-Ended Simulation Games</td>
<td>A D ○ ● ○ ○</td>
<td>Hedonic Novelty Max. (RL) ● ○ ○ ○</td>
</tr>
<tr>
<td>Togelius and Schmidhuber</td>
<td>2008</td>
<td>Learning Progress To Predict Player Fun¹</td>
<td>T I ○ ○ ○ ●</td>
<td>Learning Progress Max. ● ● ● ●</td>
</tr>
<tr>
<td>Merrick</td>
<td>2008</td>
<td>Adaptive NPCs in Dynamic Game Worlds</td>
<td>A D ○ ● ● ○ ○</td>
<td>Hedonic Novelty, Competence Max. (RL) ● ○ ○ ○</td>
</tr>
<tr>
<td>Merrick and Maher</td>
<td>2009</td>
<td>Curious Characters for Multiuser Games</td>
<td>A D ○ ● ● ○ ○</td>
<td>Hedonic Novelty, Competence Max. (RL) ● ○ ○ ○</td>
</tr>
<tr>
<td>Shaker</td>
<td>2016</td>
<td>Intrinsically Motivated RL for PCG²</td>
<td>T D ○ ○ ○ ●</td>
<td>Novelty, Surprise, Learning Progress, IMRL ✔ (RL) ● ● ● ●</td>
</tr>
<tr>
<td>Clements and Polani</td>
<td>2017</td>
<td>A Generic Utility Function for Cooperating Agents</td>
<td>A D ● ○ ○ ○</td>
<td>(Team) Empowerment Max. ● ● ● ◊</td>
</tr>
<tr>
<td>Pathak et al.</td>
<td>2017</td>
<td>Exploration by Self-Supervised Prediction</td>
<td>A D ● ○ ○ ○</td>
<td>Surprise Max. (RL) ● ● ● ●</td>
</tr>
<tr>
<td>Burda et al.</td>
<td>2019</td>
<td>Large-Scale Study of Curiosity-Driven Learning</td>
<td>A D ● ○ ○ ○</td>
<td>Surprise Max. (RL) ● ● ● ●</td>
</tr>
<tr>
<td>Burda et al.</td>
<td>2019</td>
<td>Exploration by Random Network Distillation</td>
<td>A D ● ○ ○ ○</td>
<td>Novelty Max. (RL) ● ● ● ●</td>
</tr>
</tbody>
</table>
recent contributions by Merrick and Shafi (2011), Merrick (2015) and Merrick (2016) for the same reason, but also because their motivational models are not intrinsic, as the calculation of reward depends on an extrinsically imposed goal. It thus violates agent-centricity as the most essential diagnostic of IR in our definition. The same applies for variants of novelty (Isaksen et al., 2015) or constrained novelty search (Liapis et al., 2013) as these measures are not calculated from the perspective of an agent. We notably exclude recent work on solving hard-exploration problems with the GO-EXPLORE algorithm (Ecoffet et al., 2019), as it presently only uses $e$-greedy exploration, rather than e.g. a model of curiosity (cf. Sec. 2.2.4). We do not consider Melhart et al.’s (2019) work, as it does not address the calculation of IR in artificial agents, but the prediction of human-reported IR from gameplay data.

We have finally (iii) trimmed the results further to contributions that demonstrate reasons to embrace models of IM in game AI and their application particularly well, thus adding to RQ.5 and RQ.6. Moreover, we included work that is relevant to motivating our novel models and applications in Ch. 6 and 7. We chose this non-exhaustive review strategy as the previous two steps have uncovered a highly unbalanced research landscape, with the vast majority of work employing models of IM in games, but not for games. More specifically, most of the identified work focuses on the engineering of general game-playing agents that perform well, i.e. score highly, across many different games (Togelius & Yannakakis, 2016) as a path towards artificial general intelligence. New contributions to this area are characterised by incremental improvements to the effectiveness, efficiency and simplicity of the used models, and exhaustive coverage does thus not contribute to the goals of this review. Since our contributions in this thesis leverage empowerment as formal IR, we selected two papers leveraging variations of empowerment in general game-playing. We complemented these with three recent contributions employing models of curiosity for their state-of-the-art demonstrations of the potential of IM models in general game-playing.

Our final selection comprises 11 related work items dating from 2006 to 2019. We report the type of study, the area of game AI addressed, and the details on the type and usage of IR in Tbl. 5.1. We discuss the detailed distinctions along with our findings in the next section.

We answer our research questions based on a typology of both reasons to embrace IR and models of IM in game AI, and abstract applications that leverage them. The reasons are given by (emergent) properties of IR and intrinsically motivated (IM) behaviour, derived from our working definition of IM models (Sec. 2.2.3). We extracted abstract applications of IR and IM models in game AI from our final selection of related work by considering each application a combination of these reasons and the requirements of one of the four core game AI research areas derived from Yannakakis and Togelius’ (2018, pp. 262-264) taxonomy: (game-)playing, NPCs, player behaviour and experience modelling and PCG. Both typologies focus on what has been concretely done, and not how IM models could be exploited in principle. We cover the latter through our novel models and applications in Ch. 6 and 7, and in our discussion of future work in Ch. 8.
Figure 5.1: Typologies of the benefits and applications of intrinsic reward (IR) and intrinsically motivated (IM) behaviour in videogame AI, derived from a systematic review of existing work. We conceive (abstract) applications to arise from combining: (i) (emergent) properties of intrinsic reward, intrinsically motivated behaviour and corollaries, following from our IM definition in Sec. 2.2.3; and (ii) the requirements of four core game AI research areas adopted with modifications from Yannakakis and Togelius (2018, pp. 262-264).
5.2.1.2 Findings

Fig. 5.1 summarises our findings in the form of two typologies. For the first typology we have found the same four properties of IR (R.1-4), four properties of intrinsically motivated (IM) behaviour (B.1-4) and two corollaries (C.1, C.2) as in the previous review on the use of IM models in CC (Sec. 4.2.2). For the second typology, we have identified 11 new, (abstract) applications of IR and IM models in videogame AI (A.1-11). We next motivate and describe each of these applications via brief reviews of the related work addressing them. We structure this review by the core game AI areas that each related work item can be associated with, ordered by the number of identified contributions. Within each area, related work is introduced chronologically.

(Game-)Playing (A.1-A.5)

We have argued in Sec. 5.1 that IM constitutes an important mechanism in driving human gameplay, and that games are not only autotelic by definition, but usually also intrinsically motivating by design. Any game in the strict sense, and thus the majority of commercial games, is designed to elicit the IM of human players to warrant continuous engagement. Working towards a game’s goals must yield IR, but vice-versa, acting to optimise IR is also likely to yield progress on these goals. These properties of games, in conjunction with properties of IM models, are leveraged by the earliest applications of such models to game AI, albeit usually unknowingly.

IR is subjective and sensitive to an agent’s embodiment and situatedness (R.2), i.e. it is by definition independent of extrinsic reward. Crucially though, IR can align with extrinsic reward (R.4), and intrinsically motivated behaviour can consequently yield task performance in the presence of sparse or in the absence of extrinsic reward (B.2). Assuming that in human game-playing, acting to optimise IR is likely to yield progress on game-internal goals, we suggest that the same holds true for intrinsically motivated artificial agents. Since many games try to attract different player types with different motivations, an alignment of agent-intrinsic with game-internal reward could be realised by different models of IM. Unsurprisingly, models of IM have been used to increase AI game-playing performance when extrinsic reward is sparse or unavailable, i.e. to model game-playing to win (A.2). But even if no goal alignment is present, IM-driven agents can still yield playful behaviour, and have consequently been used to model play in the absence of goals (A.5), i.e. when games are played with as toys. If agent-intrinsic reward aligns with human-intrinsic reward (C.2) to a sufficient degree, it can be leveraged to model human-like play (A.4) of or with games. The previous two applications however are rare; most existing work concentrates on designing general gameplaying agents capable of playing different games well, with the ultimate goal of superhuman performance. These contributions use IM models to realise A.2, but also leverage the domain and embodiment generality of IR (R.3) to employ the same agent across different games (A.3). Finally, some authors exploit that IR must not be directed towards specific outcomes, warranted

---

5 This case illustrates the abstract nature of these applications, in that related work often only uses them in combination: models of IM are not employed to yield agent behaviour across different games only, but jointly with e.g. A.2 to facilitate high performance across games.
by the requirement to be free of semantics (Sec. 2.2.3), in combination with the property of many IM models to induce skill and model development (B.4) to enable transfer learning from one game or game level to another (A.1).

Anthony, Polani and Nehaniv (2014) are first to embrace empowerment (Ch. 3) as an IR in general game-playing, thus addressing A.2., A.4 and A.5. They take inspiration from human players who, when approaching a new game, often manage to identify ‘good’ states with respect to the game’s goals before learning about these goals. Following the assumption that bounded rationality (Simon, 1957) incentivises structured decisions, they use the information bottleneck principle (Tishby, Pereira & Bialek, 1999) to realise an information bandwidth constraint on the action sequences in the vanilla empowerment (cf. Sec. 3.2) calculation. This ‘impoverishment’ of potential behaviours is facilitated by a hazy prediction of action consequences beyond the usual n-step lookahead. By incorporating these ‘soft horizon’ predictions, impoverishment results in a set of action sequences or ‘strategies’ that are not only expected to yield highly empowered, but also clearly distinguished states. They first evaluate this soft-horizon empowerment in a box-pushing scenario inspired by Sokoban (Imabayashi, 1982), where the intrinsically motivated agent successfully identifies strategies to free boxes from ‘traps’, or to push boxes away from doorways. They furthermore conduct experiments on a predator-and-prey game similar to Pac-Man (Namco, 1980) in which the player in a maze scores by surviving as long as possible by escaping ghosts that chase and kill them on touch. We have noted before (Sec. 3.3) that EM naturally entails death aversion, and the agent unsurprisingly escapes from the ghosts and thus scores without access to an extrinsic reward. Maximising their perceived control over the environment, the agent also realises a ‘kiting’ strategy, lettings ghosts come close enough to exactly predict and control their movements. The authors conclude that soft-horizon empowerment allows for the intuitive identification of (sub-)goals across different games, and propose to use it in general game-playing independently of extrinsic reward, or as ‘proto-heuristic’ while an agent still learns to optimise extrinsic reward.

Clements and Polani (2017) follow in Anthony, Polani and Nehaniv’s (2014) footsteps, but investigate whether EM can yield collaborative behaviour in a team of agents that aligns with a game’s goals, without knowledge of these goals (A.5). To this end, they propose to calculate the empowerment of an agent collective by treating it as superorganism, with action sequences representing the turn-wise acting of the individual agents, and action outcomes summarising their individual states. They simplify this calculation of team empowerment considerably by assuming full observability and deterministic dynamics. This determinism is warranted by limiting the opponent team to idle while the current team acts. An evaluation of this approach in simulations of Ultimate Frisbee with two teams of two players each shows that maximising team empowerment as IM realises strategies that are typical for and would contribute to the goals of this game: When not holding the disc, agents find space by moving away from the sides of the field and other players. They furthermore either intercept the disc thrown by opponents, pick it up from the ground, and pass it on to their team-mate. Their experiment thus demonstrates recognisable team-sports play behaviour induced by IM alone.
The previous insights on empowerment are of immediate relevance for our contributions, but general game-playing research at present is dominated by curiosity models of IM. To be representative, we thus include three recent contributions with state-of-the-art results. Pathak et al. (2017b) propose a model of curiosity which maximises the prediction error of an agent’s forward model as intrinsic surprise reward in RL, and is thus closely related to Schmidhuber’s (1991) early models of artificial curiosity (cf. Sec. 2.2.4). They sidestep one of the major caveats of this original approach – that agents could get attached to random sources of noise – by learning a state encoding that only distinguishes states that the agent can influence. They demonstrate several application of IM for game-playing through evaluations on Super Mario Bros. (Nintendo R&D, 1985) and ViZDoom (Kempka et al., 2016), an AI research platform based on Doom (id Software, 1993). They show that complementing extrinsic with intrinsic reward yields quicker convergence and higher performance than optimising extrinsic reward alone. They also find that an agent learns useful skills and explores much of the state space by only optimising IR. In Super Mario Bros., the agent learns how to jump over hazards and kill enemies, and thus manages to cross 30% of the first level. Finally, an agent that was pre-trained on one level using only IR achieves higher performance on previously unknown levels of the same game than an agent trained on these levels only. This serves as a quantitative proof that curiosity can yield play performance in combination with or in the absence of extrinsic reward (A.2), facilitates generalisation across different games (A.3), and supports transfer learning (A.1).

Burda, Edwards, Pathak et al. (2019) add further evidence to A.2 and A.3 through a large-scale study of the same model across 54 environments, including Super Mario Bros. (Nintendo R&D, 1985) and 48 classic titles from the Atari Game Suite (Bellemare et al., 2013). The latter comprises several notoriously hard-exploration games\(^6\) such as Montezuma’s Revenge (Utopia Software, 1984), which have proven to be very challenging for general game-playing agents that rely on extrinsic reward alone (cf. Mnih et al., 2015). In their study, Burda, Edwards, Pathak et al. (2019) note a ‘high degree of alignment between the intrinsic curiosity objective and the hand designed extrinsic rewards’ (ibid., p. 1), resulting in better performance than random agents in 75% of the Atari games. Following an increase in the batch size to improve training stability, their curious agent passes 11 levels of Super Mario Bros., learning skills such as finding secret rooms and defending bosses.

Burda, Edwards, Storkey et al. (2019) outperform Pathak et al.’s 2017 surprise reward model with a simpler and more efficient alternative. It consists of a fixed and randomly initialised target network to embed observations, and a prediction network which is given the observation and trained to predict its embedding. With repeated exposure to the same observations, the prediction network ‘distils’ the fixed target network. The error thus decreases the more familiar the observation is, and is consequently used as a novelty\(^7\) reward.

---

6 This class covers games for which extrinsic reward derived from score is either extremely sparse, or deceptive by providing misleading guidance to the overall goal (Ecoffet et al., 2016).

7 In contrast to Pathak et al. (2017b) who maximise the error in predicting future states, Burda, Edwards, Storkey et al. (2019) maximise the error in predicting the embedding of present states. They thus realise a novelty, rather than a surprise reward (cf. Grace & Maher, 2015).
They show that learning a policy which maximises only this IR through RL allows an agent to discover 15 rooms in *Montezuma’s Revenge* (Utopia Software, 1984) on average; complementing it with sparse extrinsic reward allows an agent to increase their performance and discover 19 rooms on average, occasionally passing the first level. A comparison against several baselines on five additional hard-exploration *Atari* games yields similar or better performance in five of six cases, outperforming Pathak et al.’s model in three out of six games. These results highlight the use of IM models to realise efficient game-playing in the absence of or in combination with sparse extrinsic reward (A.2), and they quantitatively support our claims in Sec. 5.1.

**Non-Player Characters (A.3-A.8)**

Non-player characters (NPCs), i.e. any character that is not controlled by a player, are a key ingredient of videogames, and contribute critically to player experience: *enemy* characters represent one of the most popular means to introduce challenges to a game (Denisova, Guckelsberger & Zendle, 2017; Denisova et al., 2020), *team-mates* or *sidekicks* can ease these challenges and familiarise a player with a game, and *neutral* characters such as *quest givers* or *conversation partners* can contribute to the believability of game worlds (cf. Warpefelt, 2016). Crucially though, NPCs also pose serious design and engineering challenges. While the previous related work has used games mostly as AI benchmarks, the following applications of IM models to steering the behaviour of NPCs exploit these models for games, i.e. for the benefit of engineers, designers and players.

This area of game AI shares three applications with (game-)playing: models of IM have been used to make NPCs exhibit playful behaviour in the absence of goals (A.5), to increase their human-likeness (A.4), and to employ them across different games (A.3) without changes to the controller. They have furthermore been exploited to create characters with individual differences using the same model (A.6). This is possible because IR is domain and embodiment general (R.3), but also subjective and sensitive to an agent’s embodiment and situatedness (R.2); characters that populate different places of a game world, and are thus exposed to different experiences, can produce different rewards and behaviours via the same model. The property R.2 is also exploited to allow NPCs to respond autonomously to unanticipated events or changes in complex, open-ended game worlds (A.8). The subjectivity of IR (R.2) warrants independence from external stakeholders, thus enabling characters to respond to situations that an AI engineer might not have anticipated beforehand, and has hence not designed an extrinsic reward for. The application A.8 also requires characters to adapt to changes in open-ended game worlds, including changes to themselves and to their abilities. This is facilitated by the potential of intrinsically motivated behaviour to yield open-ended adaptation to different domains, agent embodiments and tasks (B.3). Finally, models of IM have been embraced to increase the richness and complexity of NPC behaviour (A.7), leveraging the subjectivity of IR (R.2) combined with the potential of intrinsically motivated behaviour to induce skill and model development (B.4).

This agenda has been considerably shaped by the work of Merrick and Maher. In 2006, they highlight a discrepancy between the abilities of contemporary game AI, and the requirements of persistent virtual worlds such as
massive multiplayer online role-playing games that became increasingly popular at the time. They note that these game worlds change ‘as players create and personalise their own virtual property’ (Merrick & Maher, 2006, p. 3), but most NPCs ‘possess a fixed set of pre-programmed behaviours and lack the ability to adapt and evolve in time with their surroundings’ (ibid., p. 3). To overcome this impasse, they propose to motivate NPCs intrinsically, thus introducing application A.8. They introduce a model of curiosity which is strongly inspired by psychology and earlier work by Saunders and Gero (2004, cf. Sec. 4.2.2) to make skill development independent of the specific domain, but ‘dependent on the agent’s environment and its experiences’ (Merrick & Maher, 2006, p. 5). Their RL agents learn a policy which maximises a hedonic novelty reward calculated on events as differences of symbolic sensor states. Each event is clustered by a habituated self-organising map, and its novelty is determined as a function of the clustering error. The actual reward is obtained by transforming the novelty value through an inverse-U shaped hedonic function (Wundt, 1874; Berlyne, 1971), yielding high reward for events that are neither too novel nor too familiar. Merrick and Maher (2006) conduct a qualitative study of agent behaviour in a simple role-playing game scenario implemented in Second Life (Linden Lab, 2003). The same model is deployed on two different characters to demonstrate different emergent behaviours, realising application A.6. A ‘partner’ character progressively and autonomously evolves new behaviours in response to experiences while exploring the environment, e.g. forging weapons and mining iron. A ‘support’ character furthermore demonstrates adaptation to changes in the game: once built, another character’s house is incorporated into their exploration route.

In a later publication, Merrick and Maher (2007) draw our attention to ‘a new generation of virtual worlds’ (ibid., p. 127) within which anyone can design and modify entire games in an open-ended way. While traditional game AI such as simple reflex agents could be custom-tailored to populate these games with NPCs, Merrick and Maher note that this ‘requires development effort from game designers and, while compelling for some gamers, is too difficult or simply uninteresting for others’ (ibid., p. 129). They propose to overcome this need for manual labour by employing intrinsically motivated characters capable of responding ‘autonomously to unpredictable, open-ended changes to their environment’ (Merrick and Maher, 2007, p. 127; A.8), across any game defined within the virtual world (A.3). A major contribution of this work consists in stressing the benefits that IM models can yield for game engineers, designers and players, in terms of saving labour and opening up creative possibilities. The authors furthermore demonstrate their earlier curiosity model in a multi-agent scenario: a sheep herding game designed within Second Life (Linden Lab, 2003) in which the player must attract and keep the attention of as many intrinsically motivated sheep as possible. Crucially, each sheep can sense the player-controlled avatar but not other sheep. Qualitative observations of game-play sequences show that the sheep become attached to the player avatar and follow them into other parts of the world where they acquire new skills, e.g. using a food dispenser. Merrick and Maher also promote application A.6 by noting that, since ‘the experiences of each sheep are different, their responses to similar situations can vary, creating a cast of different characters’ (ibid., p. 133).
Merrick (2008b) adds A.7 and A.4 to the previous applications by proposing IM models as means to increase the richness and complexity as well as the human-likeness of NPC behaviour. She predicts these approaches to yield more interesting characters for the player to interact with, which can vice versa increase the believability of game worlds. She notes that curiosity as an IM may draw an agent’s attention away from half-learned tasks, i.e. the achievement of a certain event, and hence complements it with a competence motivation to maintain the ‘focus of attention for long enough to ensure that a stable behavior has emerged for completing that task’ (ibid., p. 11). The introduced competence reward is inspired by the self-determination theory interpretation of competence as an optimal challenge (Deci and Ryan, 1985; and Sec. 5.1.2), and obtained by transforming the RL agent’s q-learning error through a hedonic function. As long as the error is high, action selection and thus task performance is not stable; motivated by competence alone, an agent would thus stick to tasks that it can perform to a certain extent, but has not fully mastered yet. Merrick lets curiosity and competence motivation compete by choosing the maximum value as a motivation signal. She evaluates this combination by introducing the quantitative metrics of behavioural variety and behavioural complexity, and assesses them in a custom testbed which recreates the 2006 role-playing scenario. She finds that NPCs motivated by curiosity and competence can learn multiple, complex tasks and adapt to changes in those tasks and their environment on the fly, consequently increasing their behavioural variety. Crucially, NPCs that are motivated by both curiosity and competence are found to be more adaptable in dynamic environments than those motivated by curiosity alone.

Merrick and Maher (2009) summarise the previous studies in a book along with a strong grounding of their approach in established game AI techniques and psychological theories of motivation. They furthermore propose to use their hedonic novelty and competence reward also in a multi-option and a hierarchical RL setting to yield more complex behaviours and to speed up learning. Counterintuitively, a quantitative evaluation shows the highest behavioural variety and complexity for the vanilla, followed by the hierarchical RL agent. Merrick and Maher suggest that this indicates a trade-off between the ability of intrinsically motivated NPCs ‘to adapt quickly to changes in their environment and their ability to recall learned behaviours’ (ibid., p. 149).

All previous studies leverage IM models not to create NPCs that can win a game, e.g. to be used as opponents, but to enhance the player’s experience with believable, rich and interesting play (cf. Yannakakis & Togelius, 2018, p. 266, 268). Despite considering how well such NPCs can learn and adapt to changes in various tasks, these tasks are never evaluated against a game’s goals. We thus hold that these studies embrace IM to model play in the absence of goals (A.6), although this is not stated explicitly.

**Procedural content generation (A.9-A.11)**

PCG summarises different methods ‘for generating game content either autonomously or with only limited human input’ (ibid., p. 151). We can distinguish constructive methods through which content is generated in one pass without explicit evaluation, and generate-and-test methods that alternate between generating and evaluating content until a good result is reached.
One of the most popular examples of the latter approach is search-based PCG (Togelius et al., 2011), in which content generation is modelled as a search through a solution space (Yannakakis & Togelius, 2018, p. 157). Within this paradigm, ‘evolutionary computation has so far been the method of choice’ (Togelius et al., 2011, p. 174). We point this out, because the following applications of IR and IM models could be combined with evolutionary algorithms and thus expand the space of search-based PCG methods.

Existing work has specifically focussed on the application of IM models to efficiently explore game content spaces (A.10), leveraging that IM can give rise to exploratory behaviour (B.1). Furthermore, researchers have advocated such models to create content without game domain knowledge (A.9). This is possible because IR is domain and embodiment general (R.3). For the produced content to be meaningful to others, it is important that IR can align with extrinsic reward (R.4), e.g. in terms of expectations towards content novelty or value (C.1).

Shaker (2016) promotes the use of IM for PCG by drawing on existing formalisations of intrinsically motivated RL (e.g. Singh, Barto & Chentanez, 2005; Schmidhuber, 2010). Her contribution goes beyond IM, in that she also discusses how PCG can benefit from RL more generally, irrespective of the reward type. In the following, we only discuss proposals that leverage the combination of RL and IR specifically. Here, we exclude several of her proposals that violate our IM model definition8, which leaves us with two applications. Shaker notes that existing work on IR has been effective at modelling ‘abstract qualities such as beauty [and] novelty’ (ibid., p. 455) and consequently proposes intrinsically motivated RL as a promising candidate to generate ‘new, yet interesting and novel content’ (ibid., p. 452) – a central goal of PCG. Drawing on the potential of this framework to yield exploratory behaviour, she advocates its use for the ‘efficient exploration of the content space and higher chances of creating diverse and interesting artefacts’ (A.10) (ibid., p. 458). She envisages an RL agent that navigates the game content space by advancing individual content instances through reward-optimising actions, rather than modifying entire content populations through fitness-based selection and random modification as in evolutionary systems. She encourages combining intrinsic reward to yield novelty, and extrinsic reward such as measures of playability to warrant high quality. Shaker (ibid., p. 455) introduces a second application by pointing out that intrinsically motivated RL is applicable ‘to problems for which domain knowledge is only partially observable or expensive to obtain’, e.g. ‘when the goal is to improvise new types of games from scratch with no or minimal domain knowledge’ (A.9).

In an effort to bridge from her work on intrinsically motivated NPCs to other game AI areas, Merrick (2008b) proposes to endow game elements such as ‘buildings, weapons, furniture, and landscape’ (ibid., p. 30) with intrinsically motivated agency, thus transforming ‘virtual game worlds into adaptive virtual spaces that can evolve and change over time’ (ibid., p. 30, emphasis added). She provides several examples: ‘a weapon might develop new fighting skills, a room might learn how to trap intruders, trees might learn to repel lumberjacks’ (ibid., p. 30). We only mention this ‘agent-based

---

8 We have omitted Shaker’s (2016) use-cases of (i) experience-driven PCG and (ii) mixed initiative design-tools as the proposed RL reward relies on player or designer feedback, thus violating agent-centricity as a diagnostic of IM (Sec. 2.2.3). Tbl. 5.1 describes the remaining use-cases.
approach to PCG’ (Merrick & Maher, 2009, p. 195) as a borderline case, since the design of NPC behaviour is traditionally not considered an instance of PCG (cf. Yannakakis & Togelius, 2018, p. 151).

**Player Modelling (A.4, A.11)**

*Player modelling* concerns ‘the detection, prediction and expression of human player characteristics that are manifested through cognitive, affective and behavioral patterns while playing games’ (ibid., p. 203). In game AI, this comes down to computationally modelling ‘a player’s experience or behavior (...) based on theoretical frameworks about player experience and/or data derived from the interaction of the player with a game’ (ibid., p. 206). Given the important role of IM in human play (cf. Sec. 5.1) and the grounding of IM models in psychological theories (Sec. 2.2.4), such models are promising candidates to model human *behaviour* and *experience* in play.

Surprisingly, existing applications of IR and IM models, in the sense of our working definition in Sec. 2.2.3 are rare. Player modelling is often utilised within other core areas of game AI, e.g. to ‘improve the human-likeness and believability of any agent controller’ (ibid., p. 274). We consider the use of IM models to *model human-like play* (A.4) such a case of player *behaviour* modelling in the service of *(game-)playing* and NPCs, especially as it has been strongly grounded in observations of players (Anthony, Polani & Nehaniv, 2014) and in psychological theories of IM (Merrick, 2008b). Similarly, models of IM have been proposed for player *experience* modelling in PCG, more specifically to *predict people’s experience of fun in games* (A.11), leveraging that computational IR can correlate with human IR (C.2).

The latter application has been introduced by Togelius and Schmidhuber (2008) in a borderline case of *experience-driven PCG* (Yannakakis & Togelius, 2011). We say ‘borderline’, as instantiations of this paradigm usually use data from actual human player experience on existing content to drive the generation of new content that elicits a certain experience. Remarkably though, Togelius and Schmidhuber propose to evolve game rules based on a *model* of player experience that is *not* informed by actual human data: an agent’s *learning progress* (cf. Schmidhuber, 2010) as a predictor of people’s experience of *fun* in games (A.11). They support their approximation with psychological and game design theories such as Koster’s (2013), who argues that well-designed games start easy, but afford the player to continuously learn something new through play. They propose to use this estimate of a player’s fun as a fitness criterion in an evolutionary algorithm, which represents an *indirect use of IR* for *search-based PCG*. They do not clarify the benefits of their human-less approach specifically, but treat it as a stepping stone towards *automated game design*, e.g. to inspire human designers through innovative game prototypes, or to fine-tune a human-made game to a certain difficulty level. In their applied study, learning progress is approximated by the average score of different evolved game-playing agents. However, this does not qualify as an IR as it violates agent-centricity (cf. Sec. 2.2.3 and example on deep q-learning in Sec. 2.2.4). We thus only consider their theoretical proposal on intrinsic formalisations of learning progress as a part of related work.
5.2 Review of Intrinsic Motivation in Videogame AI

5.2.1.3 Discussion

Based on our review, we have associated reasons to embrace IR and models of IM with abstract applications of these rewards and models in the four core areas of game AI: (game-)playing, NPCs, player modelling and PCG (Fig. 5.1). Through our individual summaries for each area, we have partially answered why IR and IM models have been embraced in game AI (RQ.5) and how these have been applied so far (RQ.6). We now add to these answers by identifying and discussing usage patterns in the reviewed work.

Our first two review steps have uncovered that IM models are almost exclusively leveraged for general game-playing as a part of (game-)playing research. Related work in this popular research area treats games as benchmarks for artificial general intelligence, but does not consider how such models could advance games by benefitting engineers, designers and players. The opposite applies for all other game AI research areas: there is very little related work, and it exclusively focusses on leveraging IM models for games.

Yannakakis and Togelius (2018, p. 264 ff.) highlight that all four areas of game AI heavily rely on and influence each other. Unsurprisingly, we find a similar overlap of IM applications in related work. Most applications outside (game-)playing focus on the design of NPC behaviour, but they have largely been driven by the same group of researchers. Applications of IM models for player modelling and PCG in contrast have been advocated by different people, but they are often only touched on and lack support through applied studies. IM-driven PCG has notably only been addressed theoretically so far.

We note that strikingly many applications address some form of generality. This resonates with Togelius and Yannakakis’ (2016) appeal to consider generality not only in (game-)playing, but across all areas of game AI to advance AI for and beyond games. They formulate a research agenda towards AI that can (i) work on any game (game generality), (ii) ‘model, respond to and/or reproduce the very large variability among humans in design style, playing style, preferences and abilities’ (ibid., p. 469) (user/designer/player generality), and (iii) be applied to multiple tasks in the game design process (task generality). Related work has addressed each of these types, and in different game AI areas. Most prominently, (i) game generality is covered by leveraging IM models to model play to win with no extrinsic reward (A.2), to model play in the absence of goals (A.5), to employ the same agent across different games (A.3), to allow NPCs to respond autonomously to unanticipated events or changes in complex, open-ended game worlds (A.8), and to create game content without domain knowledge (A.9). Moreover, (ii) user/designer/player generality has been addressed by using IM models to replicate human-like play (A.4), to create NPCs with individual differences using the same AI (A.6), and to predict people’s experience of fun in games (A.11). Finally, task generality (iii) has only been explicitly promoted by Merrick and Maher (2008; 2009) when proposing a PCG application of their NPC models. More implicitly, the potential of IM models for task generality is supported by the successful use of similar models, e.g. curiosity, across different game AI areas. Our earlier observations may have suggested that IM research on generality is only conducted within (game-)playing, and in disregard of benefits for designers and players; we have
now learned that generality is also addressed in all other areas of game AI, but with a focus on benefiting games.

The diversity and limitations of IM models in related work are reminiscent of the CC case (Sec. 4.2.2). The motivational landscape is small. Most existing work focusses on some form of curiosity, realised through maximisation of a (hedonic) novelty, learning progress or surprise reward. This is complemented by a competence motivation in Merrick and Maher’s (2008; 2009) NPC work, and Shaker’s (2016) brief mention of the IMRL model for PCG. Most utilise IR directly for action selection; only Togelius and Schmidhuber (2008) propose its indirect usage as fitness in evolutionary search.

Related work considers different conceptualisations of IM, and it is thus not unexpected that several associated models do not fit all diagnostics of IM under our definition (cf. Sec. 2.2.3). Merrick and Maher’s (2006) curiosity reward for instance, used throughout their later work, is calculated from a symbolic description of sensor states. It thus relies on a closed knowledge base, which rules out freedom of semantics and embodiment sensitivity, and severely limits open-endedness. Most of the IM models referenced by Shaker (2016) conform with all our diagnostics but IMRL, which, as discussed in Sec. 2.2.4, does not warrant open-ended development. Crucially, the affected models cannot be leveraged for all game AI applications that other, fully-qualifying models have been used for. This particularly concerns applications addressing different forms of generality, as discussed and summarised earlier.

We conclude this discussion by answering RQ.7: ‘How do existing applications of IR and IM models in videogame AI and CC overlap?’ Our goal is to understand to which degree related work in game AI can be considered examples of computational game creativity (Liapis, Yannakakis & Togelius, 2014). We first consider whether game AI applications can be considered special cases of CC applications. For this purpose, we compare the (abstract) applications in our game AI typology in Fig. 5.1 to the CC typology in Fig. 4.2. We begin with cases of artefact generation and evaluation, as the most promoted examples of computational game creativity (cf. Liapis, Yannakakis & Togelius, 2014; Ventura, 2016a). We can understand the efficient exploration of game content spaces (A.10) as a means to produce novel and valuable content in PCG a special case of exploratory and transformational creativity (CC A.6, A.8) in CC, the latter when e.g. model or policy adaptation is involved. IM has been proposed as the foundation of an agent-based model of content creation within the search-based PCG paradigm (Shaker, 2016). In each step of the process, the agent assesses the novelty and potential value of content instances in the form of a (combined) reward signal, thus assessing the creativity of (partial) artefacts (CC A.1). Based on the agent-centric nature of that system, we should also consider it a model of p-creativity (CC A.2). Through A.10, IM has been applied not only to create more diverse and qualitative content, but also to do so more efficiently – i.e. to tackle the complexity of creative search (CC A.11). We furthermore consider the use of IR to predict people’s experience of fun in games (A.11) in player modelling a special case of approximating the human aesthetic judgement (CC A.5). Togelius and Schmidhuber (2008) propose to use IR as estimate of player appreciation in game content generation, thus creating artefacts that appeal to people (CC A.7). Albeit receiving little coverage in discussions of computational game creativity (cf. Zook, Riedl & Magerko,
Liapis, Yannakakis & Togelius, 2014; Moffat, 2015), we can also identify applications of IR and IM models to (game-)playing and NPCs as specialisations of CC applications. Most notably, leveraging IM models for NPCs to respond autonomously to unanticipated events or changes in complex, open-ended game worlds (A.8) echoes the CC applications to model mini-c creativity in development and adaptation (CC A.12) and to increase creative autonomy (CC A.3). Together with A.8, the creation of NPCs with individual differences from the same IM model should be considered a case of modelling p-creativity (CC A.2).

We find that related work on game AI not only specialises existing CC applications – it also advances CC by leveraging IM models for general game AI. Wiggins (2018) points out that CC shares similarities to artificial general intelligence, but notions of generality are practically rarely addressed in what is commonly considered CC (Sec. 4.2.1). Loughran and O’Neill (2018) for instance note that CC systems are usually tested on one domain only, and without further inquiry into a specific domain’s effect on the system’s performance. All facets of general game AI put forward by Togelius and Yannakakis (2016) comprise an element of novelty, e.g. in the application of the same game-playing agent to a new domain; moreover, they encompass value, as in the requirement to score highly. Without advancing this argument further, we propose that all facets of generality addressed in the reviewed related game AI work carry a strong element of creativity, which has only been addressed to a limited extend in CC research. We answer RQ.7 by concluding that all applications of IR and IM models to game AI in related work qualify as instances of computational game creativity. Research on IM in CC and game AI can thus mutually benefit each other. In the next section, we contextualise our applied contributions based on this heritage.

5.2.2 Contextualising Our Contributions

We have identified several applications of IR and IM models to game AI as special cases of similar applications in CC. These bridges have likely not been established deliberately. In this section, we account for this shortcoming and re-introduce our CC contributions in Ch. 6 and 7 through the lens of game AI. We later motivate these applications of IR and IM models to computational game creativity theoretically, and support them through experiments. This allows us to highlight hands-on applications of more generic CC research to game AI, and to pave the way for their adaptation in other CC scenarios. We show how each contribution draws on and takes the reviewed game AI work further, thus complementing the CC motivation in Sec. 4.2.3.

Our review has revealed a low diversity in the IRs and IM models used especially outside (game-)playing, with different forms of curiosity dominating the motivational landscape. We alleviate this by applying variations of empowerment and empowerment maximisation (EM, Ch. 3) to the engineering of NPC behaviour (Ch. 6) and at the crossroads of player modelling and PCG (Ch. 7), thus realising the first applications of this IR and motivational model to game AI beyond general game-playing. We carefully justify the use of empowerment and EM in game AI by considering its relationship to different
game design concepts and psychological theories of IM investigated in games user research as discussed in Sec. 5.1.1 and 5.1.2.

In Sec. 4.2.3, we have motivated the use of IM models in CC to steer the behaviour of co-creative agents in complex, open-ended interactions towards social dynamics that can benefit their interaction partners. In Ch. 6, we propose a model of social intrinsic motivation based on EM that can constrain these social dynamics to behaviours between the extremes of support and antagonism. Modern video games represent particularly complex, open-ended environments, and we consider the interaction between an NPC, a player and potentially other characters in advancing play a particularly fascinating co-creative act. We bridge from CC more generally to computational game creativity by applying our model to the design of game characters that either support or challenge the player, thus acting as companions and adversaries.

This contribution is inspired by and advances earlier work by Merrick and Maher (2006; 2007; 2008; 2009). We share their motivation to increase the believability and generality of NPCs for application in and across open-ended game-worlds. But rather than designing for human-likeness, we seek specific social behaviour between the extremes of support and antagonism. To this end, we embrace a combination of different empowerment rewards, thus complementing their use of hedonic novelty and competence. In contrast to their RL approach, we greedily maximise empowerment as a pseudo-utility; our agent can thus respond to changes in the environment, but for now does not adapt over time through learning. Crucially, Merrick and Maher’s NPCs do not perceive the player as an agent, but as a mere part of the environment. This only warrants little sensitivity to other agents, and what they consider partner-like and supportive behaviour is thus mostly incidental and contingent on the character’s embodiment. Unsurprisingly, they also do not address the modelling of adversaries. Our approach in contrast rests on the explicit modelling of the player as an agent with a stochastic policy. This crucially allows us to couple the NPC’s motivation with that of the player, yielding a social intrinsic motivation. We get a substantial increase in complexity, but gain control over the emergence of genuine social dynamics through our model’s hyperparameters. We advance related work by exploiting this added control for the modelling of different NPC personas. We also probe the inclusion of additional characters in the interaction of NPC and player, thus realising what Merrick and Maher (2009, pp. 193-194) defer to future work.

The potential of empowerment to yield supportive or antagonistic behaviour is partially supported by studies in general game-playing, e.g. by Anthony, Polani and Nehaniv (2014) as well as Clements and Polani (2017), showing that empowerment can implicitly align with a game’s goals. This is complemented by Pathak et al. (2017b), Burda, Edwards, Pathak et al. (2019) and Burda, Edwards, Storkey et al. (2019), who provide substantially more empirical evidence for such implicit goal alignment, but for curiosity as an IM model. Clements and Polani’s (2017) contribution is also relevant in that it investigates the emergence of cooperative behaviour through empowerment. Similar to Merrick and Maher though, they do not consider the state and policy of other agents in the reward calculation. In contrast to their cooperative behaviour, we model (unidirectional) support and antagonism.
In Sec. 4.2.3, we have motivated the use of IR to predict a person’s subjective experience of interactive artefacts. Such a predictive model can be used in conjunction with a generative algorithm to produce interactive artefacts that appeal to people, across different domains and independently of human feedback. Video games are arguably the most popular interactive artefacts in contemporary culture. But game designers never directly design play. They ‘can only design the rules that give rise to it. Game designers create experience, but only indirectly’ (Salen & Zimmerman, 2004, p. 168). Crucially, the same applies when game content is generated procedurally, rendering the prediction of a player’s experience of such generated content an interesting research challenge. Existing approaches are limited in their generality by relying strongly on player data and designer knowledge. In Ch. 7, we argue that this in turn curtails the potential of PCG, and propose to overcome this bottleneck with intrinsic reward-based player experience prediction.

We follow a similar agenda as Togelius and Schmidhuber (2008), but advance their work in several respects. Most importantly, they only address the use of IR for modelling player experience theoretically, but for their study refrain to a rough approximation in the form of extrinsic reward. We put their theoretical proposal into practice, but based on empowerment as an IR, rather than learning progress. We support this choice based on game design concepts (Sec. 5.1.1) and findings from games user research (Sec. 5.1.2). While their evaluation is anecdotal at best, we explore candidate player experiences that empowerment could predict in a methodologically sound, qualitative study with human players. Rather than focussing on the coarse concept of fun, we aim at predicting more fine-granular player experiences with higher accuracy. This rests on an understanding of player experience as multi-faceted and hierarchical. Similar to them, we use a function of IR to determine the fitness of content instances in an evolutionary algorithm. As noted in Sec. 5.2.1, this indirect usage of IR in an evolutionary algorithm for search-based PCG is different from Shaker’s (2016) proposal in which IR is directly optimised by an RL agent. In contrast to related work, we compare the benefits of such intrinsic reward-driven, human-less player experience modelling more carefully against other approaches to experience-driven PCG (Yannakakis & Togelius, 2011), and consider applications beyond.

We have shown that models of IM can be leveraged for computational game creativity (Liapis, Yannakakis & Togelius, 2014) by reference to three research questions. Based on a systematic review of related work, we have answered why (RQ.5) and how (RQ.6) IR and models of IM have been leveraged in four core areas of game AI. We compressed these insights into two typologies of reasons to embrace models of IM, and (abstract) applications of such models in game AI. To answer RQ.7, we eventually compared these typologies to their analogue in CC from Sec. 4.2.2, uncovering that all existing applications of IR and IM models to game AI qualify as instances of computational game creativity. Next, we complement the past two chapters as retrospective accounts of relevant research with novel models and applications leveraging IR and IM for computational game creativity.
Part III

MODEL DEVELOPMENT AND APPLICATIONS
In this chapter, we introduce coupled empowerment maximisation (CEM) as a model of social intrinsic motivation to increase the generality of artificial agents in co-creative interaction, while constraining their behaviour to either support or antagonism. We bridge between CC and videogame AI by applying our model to the development of general, believable non-player characters (NPCs) in the form of companions and adversaries. Via qualitative experiments in this domain, we partially answer the following research question:

**RQ.8** Can we use a model of intrinsic motivation to engineer general and social co-creative agents?

Our answer rests on understanding our game AI application as an instance of computational game creativity (Liapis, Yannakakis & Togelius, 2014). This chapter thus contributes to our overarching research questions by showing directly how a model of intrinsic motivation (IM) can advance game AI (RQ.2), and indirectly how it addresses core concerns of CC research (RQ.1).

We set out in Sec. 6.1 by introducing the concept of co-creativity. We describe how IM models have been used for the control of co-creative artificial agents, and highlight shortcomings of these models towards realising general, supportive and antagonistic behaviour in co-creativity. In Sec. 6.2, we show that similar shortcomings of related work exist in our application domain of videogame AI, specifically in engineering general and believable NPCs that can cope with the increasing complexity of modern games. In Sec. 6.3, we introduce a blueprint for social models of intrinsic motivation that can overcome the identified shortcomings of both, NPC AI specifically, and co-creativity more generally. In Sec. 6.4, we introduce coupled empowerment maximisation (CEM) as such a social IM model, motivated by the goal to drive the behaviour of intrinsically motivated companion and adversary NPCs. We argue for empowerment as a suitable, underlying intrinsic reward (IR) to give rise to supportive and adversarial behaviour in videogames. We introduce CEM informally, followed by a full and a simplified formalisation including pseudocode for the action policy calculation. In Sec. 6.5.1 and 6.5.2, we eventually provide a qualitative proof-of-concept for CEM to yield adaptive supportive and adversarial behaviour, respectively. To this end, we simulate the interaction of a CEM-driven NPC in a custom-made videogame testbed. We introduce the qualitative method of observational vignettes as basis for two exploratory studies. We probe the ability of CEM-driven characters to exhibit supportive and adversarial behaviour, to be general with respect to changes in their environment and embodiment, and to be sensitive, i.e. to respond with different behaviours, to such changes. We finally discuss the limitations of our studies and the model in its present formulation in Sec. 6.6.

We have published many of our ideas and findings, albeit in less detail. Our motivation in Sec. 6.1 and Sec. 6.2 expands arguments made by Guckelsberger et al. (2016), Guckelsberger, Salge and Colton (2016), and Guckelsberger, Salge
and Togelius (2018). In Sec. 6.4, we unify formalisations from each of these publications in a general and thorough account that serves as solid foundation for a discussion of future work. Our study on companion NPCs in Sec. 6.5.1 was originally published by Guckelsberger, Salge and Colton (2016), and Sec. 6.5.2 provides a more extensive account of our study on CEM-induced adversary behaviour published by Guckelsberger, Salge and Togelius (2018).

The core contribution of this chapter is CEM, a social model of IM capable of yielding supportive or adversarial, emergent behaviour in open-ended interaction. Its development entails the introduction of transfer empowerment as a novel, social variant of empowerment. We contribute to CC by highlighting additional benefits of IM models for driving the behaviour of co-creative agents, and by arguing for the importance of steering the social dynamics of such agents between support and antagonism via social models of IM. We contribute to game AI research by advancing existing arguments on the use of IM models to increasing the generality of NPC AI and to maintaining the believability of the controlled characters. Moreover, CEM represents the first motivational model based on IR to create non-neutral game characters, and the first application of empowerment to NPCs as domain of game AI. By arguing for specific interactions of a player and NPC to constitute co-creative acts, we are first to consider NPCs in the context of computational game creativity (Liapis, Yannakakis & Togelius, 2014). Finally, we contribute to AI research more generally by coining observational vignettes as a qualitative method for the study of AI behaviour. For a more detailed discussion of how the work presented in this chapter relates to existing research on models of IM in CC and videogame AI, see Sec. 4.2.3 and 5.2.2, respectively.

6.1 MOTIVATING SOCIAL CO-CREATIVE AGENTS

The concept of co-creativity is omnipresent in creativity studies and, by virtue of the tight connection between the disciplines, also addressed in much CC research. However, a commonly agreed on definition is still under debate (e.g. Kantosalo, 2019, p. 7ff.). We first clarify our understanding of co-creativity as a starting point to motivating the benefit of IM for steering the behaviour of artificial agents in co-creative interaction. We then argue more specifically how social models of IM can advance human-computer co-creativity.

We can safely state that co-creative systems are interactive CC systems (cf. ibid., pp. 10-12), instantiating systems theories of creativity (cf. Vygotsky, 1971; and Sec. 4.1.1). As proclaimed by Saunders (2012), ‘no model of creativity can be complete without an account of the interactions between individuals and their social and cultural environments’ (ibid., p. 223). The definition of co-creativity becomes more complex when considering the role of each system in this interaction. For our account, we assume that all involved systems contribute actively as agents to a joint creative process, thus acting as partners.

This chapter focuses on the specific case of human-computer co-creativity. Similar to the more general concept, there exist overlapping definitions under different labels. Davis (2013) defines it as a means to enable a computer to ‘contribute as a partner in the creative process’ (ibid., p. 9) alongside a person. While the focus here is on the creative process, Yannakakis, Liapis and
Alexopoulos (2014) use the notion of mixed-initiative co-creation to describe ‘the task of creating artifacts via the interaction of a human initiative and a computational initiative’ (Yannakakis, Liapis & Alexopoulos, 2014, p. 1, emphasis added). We adopt the following working definition of human-computer co-creativity as

‘collaborative creativity where both the human and the computer take creative responsibility for the generation of a creative artefact’. (Kantosalo et al., 2014, p. 1)

We understand creativity here as in the ‘standard definition of creativity’ (Runco and Jaeger, 2012; and Sec. 4.1.1). Kantosalo et al. (2014) borrow from existing work to define human-computer collaboration as

‘a process in which two or more agents work together to achieve shared goals’. (Terveen, 1995, p. 67)

We focus on alternating human-computer co-creativity in which the partners ‘take turns in creating a new concept satisfying the requirements of both parties’ (Kantosalo & Toivonen, 2016, p. 78, emphasis added).

Kantosalo and Toivonen (ibid.) formalise this type of co-creativity as an iterative search on a space of concepts, based on the creative systems framework (Wiggins, 2006; and Sec. 4.2.1). This formalisation allows us to intuitively recognise the opportunities in human-computer co-creativity: ‘alternating co-creation may help either party reach areas they could not have reached otherwise’ (Kantosalo & Toivonen, 2016, p. 80). Similarly, d’Inverno proposes that human-computer collaborations ‘could take the human creative into entirely unexplored territories’ (Mark d’Inverno; as quoted by Pérez y Pérez, 2018, p. 182). We elaborate two types of social dynamics by which a human collaborator can be taken into such new territories: by support and antagonism.

Supportive and antagonistic behaviour is omnipresent when people co-create in the wild. If we look at a painting class, we might observe teachers prescribing certain techniques to tackle a task, and students suggesting to each other different brushes or materials. We understand the students’ behaviour as supportive; the teacher in contrast is not outright antagonistic, but constructively challenges the student. Crucially, both types of behaviour contribute to shared goals in that they eventually allow people to overcome constraints in the creative process (cf. Stokes, 2005). Considered through the lens of the creative systems framework (Wiggins, 2006a, 2006b), such constraints exist, amongst others, on a person’s representation of the creative search space, e.g. of possible paintings, and on their strategy to traverse it, e.g. by adding paint to the canvas. Transforming these constraints is key to accessing previously unexplored territories, and considered a central mechanism of creativity (cf. Boden, 1990/2003). In suggesting different brushes or materials, the students support each other by inspiring the transformation of constraints on their conceptualisation of possible paintings, and on their techniques to realise them. By prescribing a certain technique, the teacher is antagonistic in limiting the space of possible creative trajectories, but in this course allows

---

1 A real example of more antagonistic behaviour is the sculptor Anish Kapoor’s claiming of the exclusive rights for the use of the Vantablack colour (Rogers, 2017).
for the mastery of a specific technique before approaching others, and for exploratory creativity (cf. Boden, 1990/2003) in the coverage of a smaller set of creative possibilities. While the students’ and teacher’s contributions to the joint creative process and product have in both examples been rather passive, interaction partners can equally support and challenge each other actively. In its simplest form, active support is given when one person advances creative search towards satisfying the goals of the other, e.g. when jointly working on the same painting. On the contrary, we have a case of active antagonism when one person’s contributions counteract the other’s goals, providing a challenge for the other to achieve their goals nonetheless. A teacher for instance could demonstrate a new technique on the canvas and leave it to the pupil to finish.

Support and antagonism in co-creativity can crucially be either, the vehicle to realise a separate shared goal, or the ultimate goal itself. In the latter case, co-creative interaction must – following the ‘standard definition of creativity’ (Runco and Jaeger, 2012; and Sec. 4.1.1) – yield a novel and valuable experience, with value being characterised by the quality of supportive or antagonistic interaction. But even if support or antagonism itself is the goal, there must exist a sub-goal towards which this behaviour is directed.

We advocate the design of artificial, co-creative agents, capable of exhibiting both supportive and antagonistic behaviour to take their human collaborator into new creative territories. Closely related, Kantosalo and Toivonen (2016) propose embracing pleasing and provoking agents in co-creativity. We find that most existing co-creative agents already realise supportive behaviour. Consider the Drawing Apprentice (Davis et al., 2014) for example, a system within which a person and a software agent take turns to draw on a virtual canvas. The agent receives a line input from the user, analyses and adopts the perceptual layer which they believe the user is currently in, and generates an improvised response. They can support the user by advancing the painting in a previously unanticipated direction, or by using a different drawing style which the person could adopt in an act of transformational creativity (cf. Liapis et al., 2013). Crucially though, while supportive behaviour seems to represent the modus operandi of co-creative artificial agents, antagonistic behaviour is very much absent. The Drawing Apprentice (Davis et al., 2014), for instance, could realise antagonism by revising some of the human partner’s strokes, or by limiting their colour palette. An example of a similar but implemented, antagonistic co-creative system is Adrian Ward’s digital artwork Auto-Illustrator, which ‘subverts the utilitarian spirit of commercial drawing software by turning graphic tools into autonomous agents with a will of their own’ (Ryan, 2010). Amongst others, the software reacts to the user’s input by drawing unpredictable graffiti and nonsensical words.

Kantosalo and Toivonen (2016) thus go too far by stating that provoking agents that challenge a person’s concepts in co-creativity ‘are so far non-existent’ (ibid., p. 82, emphasis added), but such systems are indeed very rare. Kantosalo and Toivonen suspect that such a provocative stance is opposed by the literature, in that co-creative agents are designed to not push their own agenda. We believe that this stance must be overcome as it, perhaps counter-intuitively, diminishes the benefits of human-computer co-creativity to people. If we want artificial agents to be taken seriously as partners in a creative activity, we require them to challenge us.
In agreement with our vision, d’Inverno demands artificial co-creators that ‘stimulate, challenge, [and] provoke us to work in new ways and to produce content that would not have been possible without the system’ (Mark d’Inverno; as quoted by Pérez y Pérez, 2018, p. 181). We value this quote as it emphasises that an artificial co-creator can be indispensable. One reason for this is given by the embodiment gap (Guckelsberger, Salge & Colton, 2017) between the human and artificial agent, arising e.g. through a different sensory or motor interface to the shared physical or virtual world. The robot Marimba player Shimon (Hoffman & Weinberg, 2010) for instance improvises in real-time to a human pianist’s performance. In contrast to their human partner though, the robot has four arms that can be moved independently on a shared rail. At least in theory, this grants Shimon access to parts of the musical space that are unreachable by the pianist alone. By virtue of their embodiment, Shimon could advance creative search in a direction that the human embodiment alone does not afford. This way, an agent’s unique embodiment can enhance support and antagonism in co-creativity.

At present, progress in CC is not only hindered by an ideological opposition to realising antagonistic co-creative agents, but also by technical limitations to unleashing their full potential. The vast majority of implemented co-creative agents should be considered extrinsically motivated in that they optimise, either implicitly2 or explicitly a pre-defined extrinsic reward landscape. Such extrinsic motivation however limits the agent’s autonomy from their designer, and in consequence to which extent they can generalise and sustain open-ended collaboration in co-creativity through support and antagonism.

Davis et al. (2017) hold that through ‘creative improvisational collaboration, a new form of distributed creativity arises that can lead to emergent, dynamic, and unexpected meaning to support creativity in new ways’ (Davis et al., 2017, p. 356; referencing Sawyer and DeZutter, 2009). When caused by the artificial co-creator, such unexpectedness can benefit the human collaborator in causing transformational creativity as argued earlier. But introduced through the human partner, it poses an insurmountable challenge to an extrinsically motivated agent: by adding to the creative process, the human co-creator can advance creative search into situations that have not been anticipated at the time of designing their artificial interaction partner – revealing an anticipation gap3. Unexpectedness then arises from a mismatch between e.g. the hard-coded situation-action rules of a reflex agent and the new situation, or the absence of extrinsic reward. Unexpectedness can also be caused by the designer failing to anticipate changes in the agent’s embodiment, altering the effect of their actions and their perception of the present situation including extrinsic reward. Consequentially, the system might suspend interaction and become ‘incapable of continuing the creative search from the concept [or artefact] provided by the other’, a scenario that Kantosalo and Toivonen (2016, p. 81) coin generative impotence. Arguably even worse, the originally sensible extrinsic reward may become deceptive, providing misleading guidance to behaviour that is counter-productive to co-creativity in the present situation.

2 We also subsume traditional reflex- and goal-based agents under this heading, as they can be defined in terms of greedily maximising a specific extrinsic reward landscape.
3 This equally applies for agents capable of adjusting their policy through a learning algorithm.
Given a sufficiently complex creative search space and a rich interface for the human collaborator to control its traversal, there is no means\(^4\) for the designers to bridge this **anticipation gap** by means of extrinsic motivation.

System designers can acknowledge this anticipation gap and try to work around it. This however necessitates a trade-off between the **generality** and **quality** of the co-creative agent’s behaviour. On the one hand, an agent could be designed for behaviours that work in many situations, but these are likely too generic to contribute to a shared goal in co-creative interaction. On the other hand, they could be tuned to high quality behaviours that only work in a few situations. Since both human and technical resources are limited, it is impossible to realise both generality and quality; using a model of extrinsic motivation, there always remains a risk for **generative impotence** or behaviour that is detrimental to co-creativity. We conclude that extrinsic motivation puts a bound on the artificial agent’s autonomy from their designer, and consequently their ability to remain responsive and contribute meaningfully to shared goals in open-ended human-computer co-creativity.

Models of **IM** can alleviate this **generality–quality** trade-off. They yield highly **general** behaviour because, following our working definition in Sec. 2.2.3, IR must be **agent-centric** and **free of semantics**; an intrinsically motivated agent will consequently always remain responsive, which rules out **generative impotence**. Moreover, being **embodiment universal**, IR does not cease, but adjusts when an agent’s embodiment changes. It is less obvious though how IM models can yield **qualitative** behaviour in co-creative interaction: while such models by definition cannot direct agent behaviour to achieving a specific separate goal, co-creativity explicitly requires collaboration on shared goals (cf. Kantosalo et al., 2014). Crucially though, an artificial agent’s IR can **align** with both the intrinsic and extrinsic rewards of a person, thus allowing for the design of an artificial agent that can creatively contribute to the human partner’s goals. Given these advantages, it is not surprising that more than one third of existing research on models of IM in CC as identified in Sec. 4.2.2 also addresses human-computer co-creativity (cf. Tbl. 4.1).

A particular challenge to generality is present in **physical** human-computer co-creation, where the possibilities for creative search are open-ended and clear-cut interaction interfaces are absent. With **Curious Whispers** (cf. Sec. 4.2.2, Saunders et al. (2010)) demonstrate that such physical co-creative interaction with a human partner can emerge from only motivating the artificial agent intrinsically. A motivation to maximise the **hedonic novelty** of simple tunes makes simple mobile robots generate their own tunes, and listen to those of others. Realising the roles of author and critic, the robots do not ‘care’ whether their interaction partners are other robots or people: Saunders, Chee and Gemeinboeck (2013) show how human participants, equipped with a synthesiser, co-create tunes with the robots. This example also demonstrates that reward alignment is a two-way street: while an appropriate choice of embodiment, situatedness and IM model can lead to the alignment of agent-intrinsic reward with human reward, potential human interaction partners

---

\(^4\) The responsiveness of the agent could be sustained if a person continuously supplied extrinsic reward at runtime that causes appropriate behaviour for the respective situation. This ‘prompting’ however would diminish the autonomy of the artificial co-creator and defy the very purpose of human-computer co-creativity. We consequently rule out this option.
can be similarly aligned: they could be told what intrinsic goals different artificial peers follow, and then matched to the system with the strongest alignment. In the case of Saunders, Chee and Gemeinboeck (2013), people’s reward expectations have been adjusted to a certain extent through the recruitment as study participants.

Still, in comparison to extrinsically motivated systems, IM allows to reduce the need to manage people’s expectations, as it by definition does not contribute to immediate, specific goals but allows for long-term skill and model development with benefits for a potentially wide range of future tasks (cf. Sec. 2.1.1 and 2.2.2). This is particularly useful when defining an extrinsic reward landscape is impossible, e.g. because the precise goals of the human co-creator are not known a priori. While the agent-centricity, freedom of semantics and embodiment universality of IR (Sec. 2.2.3) allow a system designer to narrow their anticipation gap with respect to an agent’s general responsiveness, the potential of IM for open-ended development enables this gap to be reduced with respect to sustaining the quality of behaviour.

We find that existing research on IM models for computational co-creativity (cf. Sec. 4.2.2 and Tbl. 4.1) focuses on supportive behaviour only, although antagonism is technically feasible. Crucially though, existing work exclusively relies on the implicit alignment of IR to realise such behaviour. While this technique provides benefits over extrinsic motivation for the design of more general co-creative agents, it also introduces new challenges to maintaining co-creative behaviour. We identify two central drawbacks.

Firstly, the mechanisms behind the alignment of an artificial agent’s IR with people’s intrinsic or extrinsic rewards are complex and not well understood. We illustrate this with a specific challenge. As discussed earlier, an embodiment gap between the human and artificial partners can enhance co-creative interaction by introducing an asymmetry in how both can perceive and affect each other, and allowing for exclusive access to specific areas of the creative search space. Crucially, IR is by definition sensitive to an agent’s embodiment (cf. Sec. 2.2.3), thus producing a gap. However, it is presently still unclear whether a reliable reward alignment can arise from different embodiments. We thus consider existing work prone to reward misalignment.

Secondly, the co-creative relationship induced through implicit reward alignment is spurious. From the perspective of the artificial agent, a human partner is only a latent factor in how the creative search space reacts to their actions. They do not conceive their partner’s influence on a creative search trajectory as that of a separate entity. The agent would choose a trajectory that implicitly involves a person because the outcome is more intrinsically rewarding. While these emerging relations are worthwhile investigating from CC’s cognitive perspective (cf. Sec. 4.2.1), they pose a threat from the engineering perspective which requires co-creative interaction to be sustained in order to benefit the person. But without modelling a relationship in the first place, there can be no commitment to sustaining it: if the person would advance creative search towards a situation that is less rewarding than the situation

---

5 Here, we implicitly draw on extensions of the creative systems framework (Wiggins, 2006a, 2006b) towards weak notions of embodiment (Grace & Maher, 2015) and multi-agent interaction (Linkola & Kantosalo, 2019), as well as our work on incorporating actions and action selection (Linkola, Guckelsberger & Kantosalo, 2020).
arising from the agent’s contribution alone, or if another person in the agent’s proximity would advance search in a more rewarding way, the agent will break up what only from an external viewpoint appears like a relationship to pursue the more rewarding alternatives. Such a situation is easily conceivable when implicit reward alignment is slightly off, or when the person performs a suboptimal move. We thus consider co-creativity in existing work fragile. To summarise, we have argued that support and antagonism could give partners in human-computer co-creativity access to otherwise unreachable parts of the creative search space, potentially enhanced by a difference in their embodiment. However, we found that no existing approach motivates agents to exhibit these behavioural dynamics without limiting their ability to sustain open-ended collaboration and thus co-creativity. Extrinsically motivated agents on the one hand are prone to creative impotence (Kantosalo & Toivonen, 2016) and to switching to non-collaborative behaviours that can crush co-creativity. This is because, in creating an extrinsic reward landscape, system designers cannot exhaustively anticipate changes to the creative search space and agent embodiment caused by the human interaction partner or a dynamic environment. Intrinsically motivated agents on the other hand can cope with such changes by virtue of the properties of IR, thus realising more general behaviour. Existing work enables supportive and in principle antagonistic behaviour for co-creativity by implicit alignment of agent-intrinsic reward with human-intrinsic or extrinsic reward. However, this approach is prone to reward misalignment, and the spurious co-creative relationships that emerge from it are highly fragile. Overall, these drawbacks of existing motivational approaches constrain the benefit of human-computer co-creativity to people.

In Sec. 6.3, we introduce a blueprint for social models of intrinsic motivation to overcome these drawbacks. Before that, we highlight the need for such models in our application domain of videogames, to steer the behaviour of non-player characters (NPCs) in co-creative play. This serves as an additional motivation and provides guidance for our proposal of social IM models.

6.2 General, Believable & Social Non-Player Characters (NPCs)

In this section, we motivate the need for social model of intrinsic motivation to steer the behaviour of general and believable non-player characters (NPCs) that realise the roles of companions and adversaries. We introduce the concept of NPCs, and identify shortcomings of existing NPC AI towards realising general, yet believable NPC behaviour in complex videogames. We argue that existing work on driving game characters through IM alleviates these shortcomings, but falls short of realising non-neutral behaviours such as support and antagonism. We argue that agents overcoming these shortcomings should be considered co-creative, and thus bridge to the previous section.

Non-player characters (NPCs) are ‘characters within a computer game that are controlled by the computer, rather than the player’ (Warpefelt, 2016, p. 81). They are not merely decorative, but must, through their appearance and their behaviour, portray a specific role in the game world (Warpefelt & Verhagen, 2017). They can be friends, adversaries or neutral towards the player, e.g. in the form of companions and pets, enemies and opponents, or vendors and quest-givers,
Believability

The quality of NPCs is to a large extent determined by their believability, i.e. ‘the size and nature of the cognitive gap between the character player’s experience [sic] and the character they expect’ (Lee & Heeter, 2008). Emmerich, Ring and Masuch (2018) add that believability is a relational, dynamic and emergent property, in that it captures how well an NPC fulfils a ‘players’ expectations that emerge during play (...), determined by the context the NPC appears in (...) and the role it promotes’ (ibid., p. 143). NPC believability contributes considerably to central game design and player goals: Warpefelt e.g. notes that it can increase player engagement and immersion (ibid., p. 71, emphasis added). Through an online survey, Lee and Heeter (2012) moreover find a significant and strong correlation ($r = 0.66, a = 0.01$) between players’ enjoyment when interacting with NPCs and their believability.

There is common agreement that the believability of NPCs is shaped by both, their visual appearance, and their behaviour (Loyall, 1997; Emmerich, Ring & Masuch, 2018): ‘NPCs need to not only look the part, but also behave in ways that are in line with the player’s expectations and that are believable’ (Warpefelt, 2016, p. 73). Related, Togelius et al. (2013) highlight that NPC

---

6 Our proposal extends Treanor et al.’s (2015) *AI as co-creator* game design pattern, which describes the shared construction of game content or performances through play between a human and artificial agent. While they require content generation and performance to constitute a game goal, we also allow for co-creative performance to be a separate activity that may or may not contribute to the game’s goals.
believability is not only determined by the smoothness of behaviour, but essentially about how this behaviour is being controlled. They warrant ‘player believability’ if ‘someone believes that the player controlling the character/bot is real, i.e. that a human is playing as that character instead of the character being computer-controlled’. Many determinants of believable behaviour have been proposed, but our argument in this section evolves around the four factors in Tbl. 6.1 which we have identified based on existing theoretical arguments, simulation experiments and user studies.

Firstly, an NPC must (i) realise characterhood, i.e. it must be ‘actively involved in the portrayal of its role, and (...) act in ways that are conducive to convincing the player that it is indeed in that role’ (Warpefelt, 2016, p. 33). Drawing on Dennett’s (2017, p. 287ff.) concept of personhood, Warpefelt (2016, p. 33) argues that characterhood requires an NPC to behave rationally and select actions in a way that affords the attribution of intentions via the intentional stance (Dennett, 1989). Secondly, rather than repeating the same behaviours time and again, a believable NPC should exhibit (ii) behavioural diversity. Related, Yannakakis and Hallam (2004) find that behavioural diversity can contribute to interestingness, based on the simulated evolution of opponent NPCs in a Pac-Man-like (Namco, 1980) predator-and-prey scenario. Thirdly, Lankoski and Björk (2007) identify, through analysis of a specific character, that believable NPCs should be (iii) sensitive to their surroundings and body, i.e. they must be ‘aware of their surrounding [sic] and react to changes even if the player does not directly interact with the NPC’ (Emmerich, Ring & Masuch, 2018, p. 144). This also requires an NPC to make sensible use of their embodiment in order to stay within their role. Finally, Lankoski and Björk (2007) argue that believable NPCs should follow their (iv) own agenda and take initiative. Rather than being oriented towards the player only, they should also follow separate goals. Based on a user study, Lee and Heeter (2012) find that the averaged general believability of five NPCs correlates strongly ($N = 161$, $r = 0.53$, $p = 0.000$) with players’ ability to understand the character’s goals and motivations. Emmerich, Ring and Masuch (2018, p. 144) summarise: ‘the

<table>
<thead>
<tr>
<th>Determinant</th>
<th>Reference</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Characterhood</td>
<td>Warpefelt (2016)</td>
<td>Actively and rationally portray role in a way that convinces the player.</td>
</tr>
<tr>
<td>Behavioural Diversity</td>
<td>Yannakakis and Hallam (2004)</td>
<td>Avoid the repetition of behaviour in the same or similar situation.</td>
</tr>
<tr>
<td>Sensitivity to Body and Surroundings</td>
<td>Lankoski and Björk (2007)</td>
<td>Perceive and respond to changes in own body and surroundings.</td>
</tr>
<tr>
<td>Own Agenda</td>
<td>Lankoski and Björk (2007); Lee and Heeter (2012)</td>
<td>Follow own agenda and take initiative independently of player actions.</td>
</tr>
</tbody>
</table>

Table 6.1: Determinants of believable NPC behaviour summarised from the literature.
believability and enjoyment of a character may be increased by the ability of acting self-initiated. Performing actions that are not a direct consequence of a game event and following an own agenda (personal goals and motivation) brings the NPC to life (Emmerich, Ring & Masuch, 2018, emphasis added). These determinants are partly interdependent: the realisation of adversary characterhood for instance requires an NPC to have their own agenda.

Even in simple games, creating an NPC AI that realises these four determinants of believable behaviour necessitates much repetitive engineering effort and is costly. For a new game and character, a suitable AI is typically developed from scratch and custom-tailored to the specific embodiment of the character, comprising their abilities and position in the game world. Given sufficient engineering resources, traditional NPC AI such as finite state machines, behaviour trees and scripting can indeed yield believable behaviour. Crucially though, we argue in line with (Merrick and Maher, 2009; Sec. 5.2.1) that these commonly used AI techniques are not suitable for realising believable NPC behaviour in the vast and dynamic worlds that players demand in modern games. This is because they do not generalise: a behaviour tree or script that yields sensible NPC behaviour in one situation does not necessarily produce similarly robust and believable behaviour in others. This also disqualifies them as co-creative partners in play, as it rules out creative responsibility and initiative in situations which they were not designed for.

More advanced NPC AI techniques such as supervised machine learning, extrinsically motivated RL, evolutionary algorithms and statistical planning algorithms like Monte Carlo tree search (Browne et al., 2012), n-tuple bandit (Lucas, Liu & Pérez-Liébana, 2018) or rolling horizon evolutionary algorithms (Gaina et al., 2017) can alleviate this to a certain extent. However, they eventually suffer from a more fundamental problem: they all rely on AI engineers to exactly prescribe rewards or goals during development, based on possible game states that are only revealed at runtime. Anticipating these states is complicated, as they emerge from the game mechanics in interaction with the player’s behaviour and other dynamic elements such as other NPCs. This state space may be extended even further through the use of procedural content generation (PCG), or by players contributing new game content or mechanics, as highlighted by Merrick and Maher (2007). Since engineering resources are limited, the resulting anticipation gap cannot be managed in a sufficiently complex game. Encountering a situation that has not been anticipated during development, an NPC might thus fail to realise characterhood by exhibiting behaviour that violates the player’s expectations toward their role. This may be caused, amongst others, by failure to remain sensitive to their surroundings and body. The canonical means to compensate for this shortcoming of existing AI techniques is to keep an NPC’s operational area small.

Drawing on recent advances in general game-playing AI, one might argue that more general, either friendly or adversarial behaviour could be designed by training NPCs to either maximise or minimise the player’s game achievements, such as score. However, this approach comes with high training costs, potentially little behavioural diversity, and does not equip NPCs with their own agenda separate from either supporting or challenging the player.
These shortcomings of established NPC AI techniques impinge negatively on game engineers, designers and players. More specifically, the limitation of engineering resources necessitates a trade-off between the NPC’s generality and believability. If not managed carefully, this trade-off can lead to ‘primitive, unnatural behavior’ (Emmerich, Ring & Masuch, 2018, p. 142) and thus dissatisfaction of both designers and players. More generally, it limits designers’ creative potential, and vice-versa what players can ultimately experience through videogames. Crucially though, in the future, the demands on NPC AI are likely to increase further. This is particularly emphasised by progress in PCG (cf. Smith, 2014a), with one ultimate goal being to procedurally generate entire games from scratch (cf. Cook, Colton & Gow, 2016a, 2016b). Existing NPC AI affords no means to yield general and believable NPCs in such next-generation videogames.

Existing work (cf. Sec. 5.2.1 and Tbl. 5.1) has successfully used models of IM to overcome many of these drawbacks. Similar to us, Merrick and Maher (2009) aim at designing NPC AI that can scale to complex, open-ended, dynamic game worlds. However, they motivate their approach based on properties of the specific IR functions being maximised, namely hedonic novelty and competence. We in contrast argue why models of IM more generally represent candidates to addressing this challenge, based on our working definition of such models in Sec. 2.2.3. Moreover, we address specific determinants of believable NPC behaviour, rather than the generic notion.

By virtue of agent-centricity and freedom of semantics, IR is independent of and more generic than extrinsic reward. Rather than burdening AI engineers with anticipating good extrinsic rewards across all possible game states, we can rely on IR to align with a wide range of rewards such as a game’s or player’s goals. The role of IM in human (game-)play and the possibility of such implicit reward alignment is well supported by game studies and games user research (cf. Sec. 5.1.1 and 5.1.2). With IR being free of semantics, an intrinsically motivated NPC can potentially be deployed across different games without adopting its AI to game-specific knowledge, e.g. the meaning of game tokens such as weapons or power-ups. The embodiment universality of IR allows to use the same AI on different characters, and thus to save engineering time and effort. But this property also entails embodiment sensitivity, allowing for NPCs with the same AI to exhibit different behaviours depending on their experience, their position within, and their means to perceive and affect the game world. This crucially realises the requirement of believable NPCs to be sensitive to their surroundings and body in any situation, and also increases their behavioural diversity. Moreover, IM models give NPCs their own agenda in a stronger sense; intrinsically motivated behaviour emerges from the combination of this model, an agent’s embodiment and their experience, which thus allows for behavioural initiative beyond what is anticipated at design-time. Finally, many models of IM exhibit open-ended behaviour through model and skill development. Amongst others, this affords an increase in behavioural diversity, and thus contributes to an NPC’s believability.

7 In very few cases though, these shortcoming are explicitly desired. Treanor et al. (2015) for instance propose an AI as guided game design pattern for which the player actively steers an NPC away from situations where their behaviour would be detrimental.
Models of IM can alleviate the generality–believability trade-off faced by traditional NPC AI. They allow for game-, player- (cf. Togelius & Yannakakis, 2016) and NPC-generality, the latter denoting the application of the same method across different characters without modifications. Moreover, they increase NPC believability by realising behavioural diversity, sensitivity to their surroundings and body, and a stronger notion of an own agenda. Crucially though, existing IM-driven NPC AI (Merrick and Maher, 2006, 2007, 2009; Merrick, 2008) only partially supports the realisation of characterhood as a central determinant of NPC believability. This is because existing approaches only realise neutral characters⁸; no technique so far can give rise to supportive and antagonistic behaviour for NPCs to realise the roles of companions and adversaries, while retaining a high level of generality. This curtails the benefits of this approach for game engineers, designers and players.

In the next section, we combine insights from NPC AI and human-computer co-creativity more generally in a proposal of social models of intrinsic motivation that overcomes the drawbacks in both bodies of existing work.

6.3 SOCIAL MODELS OF INTRINSIC MOTIVATION

In this section, we introduce a blueprint for social models of intrinsic motivation to overcome the drawbacks of existing work on co-creative agents and IM-driven NPC AI identified in Sec. 6.1 and 6.2, respectively. We argue that, once formalised and instantiated, such models can motivate artificial agents to enter a stable co-creative relationship with a human interaction partner, characterised by behaviour that either supports or challenges the partner’s goals. We gradually introduce the three social components of our proposal, with each component addressing challenges arising from the previous stage.

Stable supportive or antagonistic behaviour towards an interaction partner requires at minimum a reference to this partner and their goals: ‘[co-creative] systems need agency, and this involves an awareness of the human creative, their goals’ (Mark d’Inverno; as quoted by Pérez y Pérez, 2018, p. 182). Existing work in human-computer co-creativity (e.g. Saunders, Chee & Gemeinboeck, 2013) has avoided such a reference by leveraging the implicit alignment of agent-intrinsic reward with human-intrinsic or extrinsic reward. From an external perspective, the emerging behaviour can appear supportive towards the partner, albeit being entirely self-referential. However, we have found in Sec. 6.1 that this technique is prone to reward misalignment, e.g. due to an embodiment gap between the agents. Moreover, it relies on the emergence of a spurious and hence fragile relationship between the partners that can easily result in the other partner being abandoned. Social IM models address these drawbacks of implicit reward alignment by letting the agent:

**Optimise a model of the partner’s intrinsic reward:** The agent models their partner’s intrinsic reward function as a proxy to the partner’s intrinsic or extrinsic goals. This requires the agent to distinguish

---

⁸ Still, such neutral characters realise at least one type of NPC-player co-creativity, in that even neutral NPCs can respond to a player’s probing in a believable and general way, cumulating in a new and potentially sought after and thus valuable interactive experience.
their partner explicitly from the rest of their world, and to model all the partner’s components, e.g. their sensor dynamics, required to compute the model of the partner’s intrinsic reward. A model of the partner’s action policy may be necessary to calculate expectations of this reward given their actions. Using a model of their immediate action consequences, the agent selects actions that are expected to yield situations in which their partner has optimum (expected) intrinsic reward.

This component establishes a genuine social relationship, in that the agent acts towards situations that are expected to be rewarding to their partner, not themselves. We hypothesise that, for a specific IR function, partner and domain, this results in behaviour that either supports or challenges the partner. We add further support to this hypothesis throughout the rest of this chapter. Optimising a model of the partner’s intrinsic rather than extrinsic reward enables an agent to leverage a potential embodiment gap between the agent and partner in support or antagonism, while retaining generality.

At this stage, the artificial agent would only optimise the IR modelled on their partner. Crucially though, co-creativity by definition requires both parties to have creative responsibility (Kantosalo et al., 2014) and to show initiative (Yannakakis, Liapis & Alexopoulos, 2014) in interaction. If the model of the partner’s IR happens to be accurate, using it as a single source of motivation may make the interaction between agent and partner converge to a single area of the creative search space and then stagnate. Moreover, being focussed on the partner’s reward only may not allow for complex antagonistic behaviour: Kantosalo and Toivonen (2016) note that provoking agents ‘can be though [sic] of having stronger opinions, defending their viewpoints and resisting changes based on human preferences’ (ibid., p. 82); but without separating the partner’s and an agent’s own reward, distinct and non-diametric viewpoints cannot be modelled. Specific to our game AI application, we have shown that having an own agenda contributes to the believability and thus quality of NPCs. Crucially though, existing work does not distinguish between an agent’s own and their partner’s goals. We address this lack of reward separation by letting agents:

**Optimise their own intrinsic reward:** The agent not only selects actions to optimise a model of their partner’s intrinsic reward, but also to optimise an intrinsic reward that is calculated on their own components and is thus sensitive to their own embodiment.

Optimising their own IR additionally grounds the agent’s behaviour in their own perspective on the world, shaped by their specific reward function, embodiment and experience. If actions are selected based on an expected future instead of a past reward (cf. Sec. 2.2.3), a model of the partner’s policy is needed to predict their intervention and impact on this expectation.

A straight-forward way to facilitate this dual optimisation of the agent’s own and their model of the partner’s reward is to combine the expectations over both rewards in a weighted sum\(^9\). However, this technique introduces a new challenge: in some cases, the agent may select actions that result in

---

\(^9\) This is the standard procedure for motivating agents based on multiple rewards, as in the combination of intrinsic and extrinsic reward by Burda, Edwards, Storkey et al. (2019).
a situation from which they cannot affect their model of the partner’s IR, i.e. they escape operational proximity. This situation may temporarily be more rewarding for the partner or the agent themselves than any other alternative, but it can break up their relationship permanently. To prevent such social reward detachment\(^\text{10}\), we propose to let the agent:

**Optimise a social influence reward** The agent not only optimises their own and the model of the partner’s intrinsic reward, but also a reward measuring their actual or potential influence on the partner.

Optimising such a social influence reward reinforces social attachment, and ensures that the agent can also in the future maintain or even increase their influence on the partner’s modelled IR. From the partner’s perspective, this ascertains that the artificial agent is not going to abandon them even if the partner decides to endure temporary reward bottlenecks, which are characteristic of many co-creativity and videogame scenarios.

In summary, social models of IM require an agent to explicitly model their interaction partner and to select actions which optimise the partner’s expected IR. Moreover, the agent must optimise their own IR, as well as an intrinsic social influence reward that prevents social detachment. In combination, these three components can alleviate the quality–generality trade-off in existing motivational models for co-creative agents. Social IM models allow agents to establish a more stable social relationship, which is characterised by supportive or antagonistic behaviour and benefits from the agent being sensitive to their own embodiment, perspective and goals. By relying on IR throughout, such models remain highly general with respect to the domain, the partner, and the agent’s embodiment.

This proposal represents a blueprint that is underspecified in various ways, and from which specific social IM models must be instantiated. To realise the first component (i), we must select an appropriate model of IR that approximates the partner’s actual IR well, and works as a generic proxy to their extrinsic goals. The calculation of some rewards requires a model of the partner’s policy. We can assume this policy to be either fully determined by the optimisation of the modelled IR (on-policy), to be independent (off-policy), or anything in-between\(^\text{11}\). To realise the second component (ii), we must decide on the agent’s own IR function. This could be informed by what type of reward would make the interaction with their partner more worthwhile, e.g. through establishing a particular opposition for sophisticated antagonism. To realise our third component (iii), we must decide on what type of social influence to optimise for through IR. To maintain operational proximity, the agent should at least influence one of the partner’s components that contribute to the model of the partner’s IR. Finally (iv), we must define how these three individual rewards are combined in action selection. The framework is not restricted to a

\(^{10}\) Social reward detachment is an artefact of optimising the expected rewards of multiple agents. This differs from the reward detachment introduced by Ecoffet et al. (2019) which is caused by a single agent’s IR being adaptive (cf. Sec. 2.2.3) and thus wearing off over time.

\(^{11}\) We borrow this notion from RL, where an off-policy approach distinguishes between a policy to generate behaviour, and another to be optimised (Sutton & Barto, 2018, p. 110). We use the notion to highlight the discrepancy between a policy used to model another agent’s behaviour, and a separate, hypothetical policy that optimises the assumed IR.
specific number of reward functions, as long as at least one function can be matched to each of the three components.

In the next section, we instantiate the motivational principle of coupled empowerment maximisation (CEM) from this blueprint to drive the behaviour of general, believable, and social NPCs.

6.4 COUPLED EMPOWERMENT MAXIMISATION

In this section, we introduce coupled empowerment maximisation (CEM), a social model of IM to drive the behaviour of general and believable NPCs that either support or challenge the player, allowing them to realise full characterhood in the roles of companions or adversaries, respectively. We first argue why empowerment, as the basis of our model, is a suitable IR to motivate supportive and antagonistic behaviour in videogames. We then introduce CEM informally and present both a general and simplified formalisation, complemented by pseudocode. We briefly discuss related work in the end.

6.4.1 Empowerment and Game Progress

We have argued in Sec. 5.1.1 that, to warrant continuous engagement, commercial videogames are designed to be intrinsically motivating for as many players as possible. The players in turn are motivated by a variety of IRs, as suggested by studies in games user research (cf. Sec. 5.1.2). The first step in instantiating a social model of IM is to decide on the three types of IR that form its basis. We have previously hypothesised that, for a specific model of the partner’s IR, the actual partner and the domain, social models of IM can yield supportive or antagonistic behaviour. For this to work in videogames, the model of the partner’s IR must (i) correspond to an IR that many players pursue, but at the same time (ii) align with the goals provided by a wide range of videogames, ideally from different genres. While this model of the partner’s IR is certainly pivotal for the emerging behavioural dynamics, the type of IR given to the agent itself can shape these dynamics further, and the type of social influence reward is important to keep the relationship between agent and partner stable over time.

Our choice of IR is based on a crucial observation: progress towards a game’s goals often goes along with the player maintaining or increasing their options and influence. We illustrate this based on classic titles that have considerably influenced their respective genre. Real-time strategy games such as StarCraft (Blizzard Entertainment, 1998) for example often require the player to collect different kinds of resources. These resources allow players to afford new buildings and upgrades, scale up the supply chain, build military units and expand the player’s sphere of influence. Throughout all these steps, the player gains more degrees of freedom. First-person shooters such as Doom (id Software, 1993) provide the player with additional power-ups and weapons in the progress of the game. Crucially, collecting a new weapon not only gives the player another, but also a more efficient means to cope with their enemies. To retain these options and influence, a player must spend their ammunition wisely and look out for replacements. Eliminating their enemies
gives the player access to new areas of the game, which again corresponds to a wider range of options. This connection crosses genre boundaries, in that e.g. collecting inventory items in role-playing games such as *Diablo* (Blizzard North, 1997), or hitting checkpoints in a racing game such as the classic arcade title *Out Run* (Sega, 1986) either directly contributes to a game’s goals, or puts the player in a better position to achieve those goals.

Crucially, *empowerment* (cf. Klyubin, Polani and Nehaniv, 2005; Klyubin, Polani and Nehaniv, 2005; Salge, Glackin and Polani, 2014; and Ch. 3) can quantify a player’s perceivable, reliable and potential options and influence in an IR. Existing studies in general game-playing (cf. Sec. 5.2.1) by Anthony, Polani and Nehaniv (2014) and Clements and Polani (2017) support the notion that empowerment can align with a game’s goals, in that their empowerment maximising agents realise several game goals without encoding these explicitly. Based on the previous observations, we hypothesise that empowerment aligns with the goals in a wide range of videogames across different genres.

Moreover, empowerment overlaps with several IRs that games user studies have linked to human gameplay motivation (cf. Sec. 5.1.2), most prominently *autonomy* and *competence* as part of self-determination theory (Ryan & Deci, 2000b; Ryan, Rigby & Przybylski, 2006). Empowerment relates to autonomy in that it is sensitive to a player’s options in a certain game state; but it also quantifies how reliably a player can influence the game world through their actions, and thus relates to competence through the concept of effectance (White, 1959; Harter, 1978; Malone, 1981; Klimmt, Hartmann & Frey, 2007).

Summarised, we hypothesise an alignment between empowerment, a formal IR quantifying an agent’s potential and perceivable options and influence (Sec. 3.2), and the goals provided by a wide range of videogames. Moreover, we note a close relationship between EM and the IMs that are thought to shape human gameplay. We consequently choose empowerment as the underlying IR for the social model of IM to be defined next: coupled empowerment maximisation (CEM).

### 6.4.2 Coupled Empowerment Maximisation: Intuition

Our goal is to realise general and believable NPCs. The previous argument supports the claim that empowerment maximisation (EM) can yield high generality in game-playing. However, we yet have to clarify why it also warrants NPC believability (Sec. 6.2 and Tbl. 6.1), especially in terms of allowing them to support or challenge the player as companions or adversaries and thus realise characterhood. To see this, recall the previously observed alignment between a player’s options and influence, and their progress in a game: we consequently expect a game to be more challenging if it was harder for us to maintain and extend our options and influence. This is supported by a study with human players conducted by Guckelsberger et al. (2017) and discussed extensively in Ch. 7, which suggests that decreasing a player’s empowerment in a game is linked to an increase in their perceived challenge (cf. Denisova, Guckelsberger & Zendle, 2017; Denisova et al., 2020).
6.4 Coupled Empowerment Maximisation (CEM)

From this follows the core idea of CEM: we drive an NPC’s action-selection to not only maximise their own empowerment, but also to either maximise or minimise the player’s empowerment, both to either maximise or minimise the player’s empowerment. We hypothesise that a character minimising the player’s empowerment would be perceived as antagonistic, and a character maximising it as supportive. To further this intuition, consider the following thought experiment about our previously used game examples. How would you perceive an NPC in ...

- ... Starcraft (Blizzard Entertainment, 1998) who sent units to knock out your power supply chain, compared to one that transferred resources allowing you to build more units?

- ... Doom (id Software, 1993) that attacked and pushed you back, compared to one that strengthened your defence?

- ... Diablo (Blizzard North, 1997) who stole part of your inventory, compared to one that gave you a powerful artefact as a gift?

- ... Out Run (Sega, 1986) who rammed your car and slowed you down, compared to one making space for you to take over?

Across all these fictional scenarios, the NPC has either decreased or increased the player’s options and influence, which can be quantified by empowerment.

We provide an informal account of CEM by describing how it specifies each component of social IM models, following the blueprint in Sec. 6.3. Firstly (i), a CEM-driven NPC chooses actions that either minimise or maximise a model of the player’s empowerment as an approximation of the actual IR driving human gameplay, and as a generic proxy to the extrinsic goals in a wide range of games. Being a measure of the player’s potential and perceivable influence on the game world, calculating player empowerment requires a model of their sensorimotor dynamics, as well as a model of their policy. CEM follows an off-policy approach in that the player’s policy is not assumed to be determined from maximising their empowerment alone; rather, the model concedes that the policy may also optimise other rewards, or be biased by the player’s relationship towards the NPC, as detailed later. Secondly (ii), a CEM-driven NPC selects actions to maximise their own empowerment, and thus – given successful reward alignment – to implicitly act towards a game’s goals. By following the same reward type as assumed for the player, the NPC can match or oppose the player’s goals, depending on whether they minimise or maximise the player’s model of IR. In either case, the NPC follows their own agenda as empowerment is shaped by their distinct embodiment and situatedness. Thirdly (iii), a CEM-driven NPC maximises transfer empowerment, a social influence reward that we have introduced specifically for the development of CEM, and that measures the NPC’s potential – not actual – influence on the player’s future perceptions. This is a suitable social influence reward, as the player’s empowerment is

\[ \text{Admittedly, the notion ‘coupled empowerment maximisation’ can be misleading: while the NPC always maximises their own and transfer empowerment, they may also minimise their partner’s empowerment. Formally though, this is always defined as maximising a coupled reward, although with potentially negative weights.} \]
sensitive to their future perceptions. Thus, even if the NPC cannot directly affect the player’s empowerment at the current point in time, maximising this reward keeps them in operational proximity, i.e. in game states where they can likely affect the player’s empowerment in the future. We demonstrate later that this can, but does not have to coincide with spatial proximity. The CEM policy combines these three reward signals in a weighted sum. In selecting the next action to perform, the NPC thus trades off the action’s impact on the player’s and their own potential and perceivable influence, as well as their operational proximity to the player. The weights represent hyperparameters of the model and a means to shape the emerging behavioural dynamics further, as demonstrated in our experiments in Sec. 6.5.

6.4.3 Coupled Empowerment Maximisation: Formal Definition

CEM can be thought of as an extension of EM to the multi-agent case, and we consequently follow a similar structure as in the original account in Ch. 3. We first extend the perception-action (PA)-loop and the agent’s generative model to model the multi-agent interaction from an objective, external and from a subjective, agent-centric perspective, respectively. The complexity of the generative model hereby increases substantially compared to the single-agent case, especially with respect to the number of involved parameters, and the difficulty in estimating them. We first formalise the different empowerment variants in CEM and the action-value function calculated on this generative model thoroughly to highlight potential avenues for future work in Ch. 8. We then introduce assumptions to simplify the formalism for our experiments.

6.4.3.1 Multi-Agent PA-Loop and Generative Model

Similar to EM, CEM as social IM model assumes an agent-centric perspective, and we thus also distinguish between the objective world that an agent is embedded in, and their beliefs about that world. For CEM, we extend both perspectives, represented by the PA-loop and generative model, to capture the interaction of the NPC with the player. We also model an ‘other’ character in interaction as a placeholder for agents that the NPC is not directly coupled with, but which can intervene in their interaction with the player.

We formalise two versions of the multi-agent PA-loop as causal bayesian networks (BNs) (cf. Appx. B). For our more general account of CEM, we include the NPC’s memory as the basis for inference from experience, with the graph structure shown in Fig. 6.1a. As in the single-agent version, we assume that the game world originates at time $t = 0$, which can be conceived e.g. as the start of the present game level. To accommodate multiple agents, we decompose the rest of the world from the original model into two more agents, player and ‘other’, that jointly contribute to the shared world. We assume that the game is discrete in time and space, and that the different characters interact in turn-wise order\(^\text{13}\). We distinguish consecutive interaction cycles, with each cycle initiated by the NPC, followed by the player, and concluded by the ‘other’ agent. Each character is represented by two random

\(^{13}\) This assumption is not very restrictive, as the individual characters may also idle.
variables modelling their sensor and actuator, associated with their owner through indices $C, P$ and $O$ for NPC, player and ‘other’, respectively.

We introduce helper variables to conveniently deconstruct the multi-agent interaction. We denote the length of an interaction cycle with $t = 3$, and the time when the NPC, player and ‘other’ character act next by $t_C, t_P$ and $t_O$. We can thus consider the interaction of characters relative to each other, and, by using $t$ as an offset, reference random variables across interaction cycles. Our fixed interaction order dictates that $t_P = t_C + 1$ and $t_O = t_C + 2$. To specify the loop, we define the following random variables and state spaces:

- **NPC, player and ‘other’ sensor** $S^C, S^P, S^O$ with state spaces $S^C, S^P, S^O$
- **NPC, player and ‘other’ actuator** $A^C, A^P, A^O$ with state spaces $A^C, A^P, A^O$
- **NPC memory** $M^C$ with state space $M^C$
- The rest of the system $R$ with state space $\mathcal{R}$

The memory state space is defined in analogy to Eq. 3.1 but with memorised actions and perceptions offset by $\tau$. We complete the multi-agent PA-loop specification by defining the causal dependencies between these random variables via the following interventional distributions (cf. Appx. A):

- **Sensor dynamics** $p(s_{t_C}^C | r_{t_C}), p(s_{t_P}^P | r_{t_P}), p(s_{t_O}^O | r_{t_O})$
- **Memory dynamics** $p(m_{t_C}^C | s_{t_C}, m_{t_C-\tau}, a_{t_C-\tau}^C)$, for the NPC only
- **Initial memory dynamics** $p(m_0^C | s_0^C)$, for the NPC only
- **Action policies** $p(a_{t_C}^C | m_{t_C}^C), p(a_{t_P}^P | s_{t_P}^P), p(a_{t_O}^O | s_{t_O}^O)$
- **Environment dynamics** $p(r_{t_C+1} | a_{t_C}^C, r_{t_C}), p(r_{t_P+1} | a_{t_P}^P, r_{t_P}), p(r_{t_O+1} | a_{t_O}^O, r_{t_O})$
- **Initial environment state** $p(r_0)$

---

**Figure 6.1:** Multi-agent perception-action loop with and without NPC memory, unrolled in time from the initial state onwards. The random variables $S, A, M$ and $R$ represent the agents’ sensor, actuator and memory as well as the rest of the system. The loop shows the first interaction cycle, with NPC (C), player (P) and ‘other’ agent (O) acting in turn-wise order.
The multi-agent PA-loop is heterogeneous in that the player and ‘other’ character are not assumed to have memory, and thus decide on their actions based on the present perception alone.

We complement this objective view with a multi-agent generative model encapsulated in the NPC’s policy, and representing their subjective view on the interaction with the player and ‘other’ character. This model is not only used for the calculation of the NPC’s own empowerment, but also the player’s empowerment and the NPC-player transfer empowerment. This requires the NPC to (i) infer latent variables such as the state of the environment and the parameters of their own, the player’s and the ‘other’ character’s dynamics. Via these inferred quantities, they can then (ii) generate predictions of how the player’s actions affect their future perceptions, and how the NPC’s actions affect both their and the player’s future sensor states as the basis for the coupled empowerment calculation. To this end, the model relates parameters and latent variables as ‘generative causes’ to the NPC’s sensor values.

We show the graph structure of the multi-agent generative model in Fig. 6.2. The causal BN comprises the following random variables:

- **NPC**, player and ‘other’ sensor $\hat{S}_C, \hat{S}_P, \hat{S}_O$ with state spaces $\hat{S}_C, \hat{S}_P, \hat{S}_O$
- **NPC**, player and ‘other’ actuator $\hat{A}_C, \hat{A}_P, \hat{A}_O$ with state spaces $\hat{A}_C, \hat{A}_P, \hat{A}_O$
- The rest of the system $\hat{R}$ with state space $\hat{R}$

---

**Figure 6.2:** Generative model of the multi-agent perception-action-loop, showing one interaction cycle initiated by the NPC (C), and followed by the player (P) and ‘other’ agent (O). (Hyper-)parameters are shown in grey. An edge connecting one to $n$ nodes (e.g. $\Theta_{C,P}^1 \rightarrow \hat{S}_C^0, \hat{S}_P^0, \ldots$) corresponds to $n$ edges from that node to each of its children ($\Theta_{C,P}^1 \rightarrow \hat{S}_C^0, \Theta_{C,P}^1 \rightarrow \hat{S}_P^0, \ldots$).
As in the case of EM (Sec. 3.2), we use a hat to denote variables that are assumed by the agent internally, e.g., $\hat{A}$ and that model variables in the PA-loop which have not been resolved yet or cannot be accessed directly, e.g., $\hat{R}$. We assume the actuator and sensor state spaces for each agent to match the state space of the modelled, objective variable, i.e., $\hat{S}_c^C = S_c^C, \hat{S}_p^O = S_p^O, \hat{S}_r^O = S_r^O$ and $\hat{A}_c^C = A_c^C, \hat{A}_p^O = A_p^O, \hat{A}_o^O = A_o^O$. Moreover, we assume that $\hat{R} = R$, i.e. the NPC knows the possible states of the external, shared world.

The model is considerably more complex than for single-agent EM, in that we distinguish three sets of continuous (hyper-) parameters, each representing parameters and the same number of hyperparameters $X = (\Xi^0, \Xi^1, \Xi^2, \Xi^3)$

- The initial state of the world\(^{14}\) as well as the NPC’s sensor dynamics, environment dynamics and policy $\Theta = (\Theta^0, \Theta^1, \Theta^2, \Theta^3)$, determined by hyperparameters $\Xi = (\Xi^0, \Xi^1, \Xi^2, \Xi^3)$
- The player’s sensor dynamics, environment dynamics and policy $\Phi = (\Phi^1, \Phi^2, \Phi^3)$, determined by hyperparameters $\Sigma = (\Sigma^1, \Sigma^2, \Sigma^3)$
- The ‘other’ character’s sensor dynamics, environment dynamics and policy $\Psi = (\Psi^1, \Psi^2, \Psi^3)$, with hyperparameters $\Omega = (\Omega^1, \Omega^2, \Omega^3)$

CEM necessitates another layer of complexity. To accurately model the player and NPC-player transfer empowerment, the NPC must switch between their own and the player’s perspective. Consequentially, the generative model must not only encode the NPC’s beliefs in the player’s sensor and environment dynamics, but also the NPC’s beliefs in the player’s beliefs in the player’s dynamics. Moreover, while neither the NPC nor the player require a model of their own policy for the calculation of empowerment, they require a model of their peers’ policies to predict how these peers could intervene in their own sensory futures. We represent this perspective switching by indexing individual parameters with $C, P$ and $O$. The generic calculation of coupled empowerment thus involves the following 17 parameters and the same number of hyperparameters:

- The NPC’s beliefs in the initial environment state and their own sensor and environment dynamics $\theta_C = (\theta_C^0, \theta_C^1, \theta_C^2)$, with $\Xi_C = (\Xi_C^0, \Xi_C^1, \Xi_C^2)$
- The NPC’s beliefs in the player’s sensor dynamics, environment dynamics and policy $\phi_C = (\phi_C^1, \phi_C^2, \phi_C^3)$, with $\Sigma_C = (\Sigma_C^1, \Sigma_C^2, \Sigma_C^3)$
- The NPC’s beliefs in the ‘other’ character’s sensor dynamics, environment dynamics and policy $\psi_C = (\psi_C^1, \psi_C^2, \psi_C^3)$, with $\Omega_C = (\Omega_C^1, \Omega_C^2, \Omega_C^3)$
- The player’s beliefs in the NPC’s sensor dynamics, environment dynamics and policy $\theta_P = (\theta_P^1, \theta_P^2, \theta_P^3)$, with $\Xi_P = (\Xi_P^1, \Xi_P^2, \Xi_P^3)$
- The player’s beliefs in their own sensor and environment dynamics $\phi_P = (\phi_P^1, \phi_P^2)$, with $\Sigma_P = (\Sigma_P^1, \Sigma_P^2)$
- The player’s beliefs in the ‘other’ character’s sensor dynamics, environment dynamics and policy $\psi_P = (\psi_P^1, \psi_P^2, \psi_P^3)$, with $\Omega_P = (\Omega_P^1, \Omega_P^2, \Omega_P^3)$

\(^{14}\) Note that, in contrast to the formalisation of EM in Sec. 3.2, the initial state of the world is now parametrised by parameter $\phi^0$ and hyperparameter $\xi^0$, rather than $\phi$ and $\xi$.\(^{15}\)
Despite modelling the player’s beliefs by proxy, all parameters are owned by the NPC. This constitutes a cognitive theory of mind (Premack & Woodruff, 1978) insofar as the coupled empowerment maximising agent models the dynamics, policies and beliefs of the other agents they interact with. As in the formalisation of EM, we assume the hyperparameters to be fixed to 
\[ \xi = (\xi^0_{C,P}, \xi^1_{C,P}, \xi^2_{C,P}, \xi^3_{C,P}) \], \( \sigma = (\sigma^1_{C,P}, \sigma^2_{C,P}, \xi^3_{C,P}) \) and \( \omega = (\omega^1_{C,P}, \omega^2_{C,P}, \omega^3_{C,P}) \).

We complete the generative model definition by specifying the causal dependencies within. These distributions represent the NPC’s models of the true dynamics in the PA-loop, and are hence denoted by \( q \) rather than \( p \):

- Action probabilities \( q(a^t_C) \), for the NPC only
- Initial environment state model \( q(\hat{r}_0; \theta^0) \), for the NPC only
- Sensor dynamics model \( q(s^0_{t_C} | \hat{r}_t; \theta^1) \), \( q(s^0_{t_P} | \hat{r}_t; \phi^1) \), \( q(s^0_{t_0} | \hat{r}_t; \psi^1) \)
- Environment dynamics model \( q(\hat{r}_{t_C+1} | \hat{a}_{t_C}, \hat{r}_t; \theta^2) \), \( q(\hat{r}_{t_P+1} | \hat{a}_{t_P}, \hat{r}_t; \phi^2) \), \( q(\hat{r}_{t_0+1} | \hat{a}_{t_0}, \hat{r}_t; \psi^2) \)
- Policy model \( q(\hat{a}_{t_C} | s^C_{t_C}; \theta^3) \), \( q(\hat{a}_{t_P} | s^P_{t_P}; \phi^3) \), \( q(\hat{a}_{t_0} | s^0_{t_0}; \psi^3) \)
- Belief priors \( q(\theta^i; \xi^i) \forall i \in [0,3] \), \( q(\phi^i; \sigma^i) \), \( q(\psi^i; \omega^i) \forall i \in [0,3] \)
- Belief hyperpriors \( q(\xi^i) \forall i \in [0,3] \), \( q(\sigma^i) \), \( q(\omega^i) \forall i \in [1,3] \)

For the calculation of coupled empowerment, most of these dynamic models are considered from different perspectives, as detailed earlier. The causal dependency \( q(\hat{a}_{t_C} | s^C_{t_C}; \theta^3) \) is not reflected in the model topology (Fig. 6.2)\(^{15}\).

We assume that the NPC can in principle infer the latent environment state and all parameters based on memorised sensorimotor experience and the hyperpriors. However, working out the inference process is subject to future work. For now, we assume the following posterior factor to be given:

\[
q(\hat{r}_t, \theta, \phi, \psi | m_i; \xi, \sigma, \omega) = \sum_{\hat{r}_{<t}} q(\hat{r}_{<t}, \theta, \phi, \psi | m_i; \xi, \sigma, \omega) \quad (6.1)
\]

The latent environment states prior to \( t \) have been marginalised out here as they are not required for the calculation of coupled empowerment. We use variations of this posterior factor where in addition a subset of the parameters \( \theta, \phi, \psi \) has been implicitly marginalised out.

### 6.4.3.2 Generic Coupled Empowerment Maximisation

Coupled empowerment combines a model of the player’s empowerment as a proxy to their goals, the NPC’s own empowerment, and the NPC-player transfer empowerment to maintain operational proximity. These IRs all measure the potential impact of action sequences of length \( n \) on future sensor states. In contrast to vanilla EM though, CEM requires us to account for the other

---

\(^{15}\) At this point, the network cannot (i) encode these NPC beliefs about the player’s beliefs about the NPC’s policy and at the same time (ii) indicate that the NPC, from their own perspective, freely chooses their actions. This could be overcome by conditioning edges on the different perspectives. However, this would complicate the model further, and we thus omit it.
agents interleaving this course of action. We consequently consider how *interleaved action sequences* $\hat{a}_t^\tau = (\hat{a}_t, \hat{a}_{t+\tau}, \ldots, \hat{a}_{t+(n-1)\tau})$ impact the NPC’s sensor state $n$, or the player’s sensor state $n-1$ interaction cycles ahead. We first formalise each empowerment type as a state-dependent quantity, following by their expected value to be used in the CEM action-value function.

The state-dependent NPC empowerment measures how much potential influence the NPC has in a specific latent state of the environment on their future sensor state $n$ interaction cycles ahead. We formalise it by extending the vanilla empowerment in Eq. 3.9 to the multi-agent case:

$$e^C(r_{tC}; \theta_C, \phi_C, \psi_C) = \max_{q(a_{tC}^n)} I(\hat{A}_{tC}^C \rightarrow \hat{S}_{tC+n|T}|r_{tC}; \theta_C, \phi_C, \psi_C)$$

$$= \max_{q(a_{tC}^n)} \sum_{a_{tC}^n, s_{tC+n|T}} q(a_{tC}^n) q(s_{tC+n|T}|a_{tC}^n, \hat{r}_{tC}; \theta_C, \phi_C, \psi_C) \log \frac{q(s_{tC+n|T}|a_{tC}^n, \hat{r}_{tC}; \theta_C, \phi_C, \psi_C)}{q(s_{tC+n|T}|\hat{r}_{tC}; \theta_C, \phi_C, \psi_C) q(a_{tC}^n)}$$

The NPC’s actions $\hat{a}_{tC}^n$ are assumed to be executed across $n$ consecutive interaction cycles, starting at time $t_C$, and offset by the cycle length $\tau$.

This channel capacity (Appx. C) is calculated on an *n-step predictive factor*, quantifying the probabilities of future NPC sensor states $n$ interaction cycles ahead. If the NPC was alone, these would be determined by their actions’ impact on the latent environment states, and the environment’s own dynamics. In the multi-agent scenario however, we must also account for the other characters’ potential impact on the shared environment. The factor is thus parametrised by the NPC’s beliefs in their own, the player’s and the ‘other’ character’s dynamics:

$$q(s_{tC+n|T}|\hat{a}_{tC}^n, \hat{r}_{tC}; \theta_C, \phi_C, \psi_C) = \sum_{\hat{r}_{tC+n|T}} q(s_{tC+n|T}|\hat{r}_{tC+n|T}; \theta_C^1) q(\hat{r}_{tC+n|T}|\hat{a}_{tC}^n, \hat{r}_{tC}; \theta_C, \phi_C, \psi_C)$$

The last term captures the impact of the NPC’s actions on the environment exactly $n$ interaction cycles ahead. We define these *n-step cyclic dynamics* recursively based on the $\tau$-step dynamics that bridge a single interaction cycle:

$$q(\hat{r}_{t+k\tau}|\hat{a}_t, \hat{r}_t; \theta, \phi, \psi) = \sum_{\hat{r}_{t+k\tau}} q(\hat{r}_{t+k\tau}|\hat{a}_{t+(k-1)\tau}, \hat{r}_{t+(k-1)\tau}; \theta, \phi, \psi) q(\hat{r}_{t+(k-1)\tau}|\hat{a}_{t-(k-1)\tau}, \hat{r}_t; \theta, \phi, \psi)$$

This definition is independent of the specific agent considered, and used for different empowerment variants. We thus omit agent-specific indices.

The calculation of the $\tau$-step cyclic dynamics in contrast is sensitive to the specific agent. It captures the influence of each agent in a single interaction cycle
on the latent environment state, and thus relies on their relative positions. For the NPC, we have:

$$ q(\hat{r}_{t+c} | \hat{a}_{t+c}, \hat{r}_{t+c}; \theta_C, \phi_C, \psi_C) = \sum_{\hat{r}_{t+c} \hat{a}_{t+c}} q(\hat{r}_{t+c} | \hat{r}_{t+0}; \psi_C) q(\hat{r}_{t+0} | \hat{a}_{t+c}; \phi_C) 
$$

$$ q(\hat{a}_{t+c} | \hat{r}_{t+c}; \theta_C^2) $$

(6.6)

The last term is the NPC’s model of their environment dynamics, and the other two encode the player’s ‘other’ agent’s impact on the environment:

$$ q(\hat{r}_{t+1} | \hat{r}_{t}; \phi^2) = \sum_{\hat{a}_{t} \hat{r}_{t}} q(\hat{r}_{t+1} | \hat{a}_{t}, \hat{r}_{t}; \phi^2) q(\hat{a}_{t} | \hat{a}_{t}; \phi^3) q(\hat{a}_{t}; \phi^1) $$

(6.7)

$$ q(\hat{r}_{t+1} | \hat{r}_{t}; \psi^2) = \sum_{\hat{a}_{t+1} \hat{r}_{t+1}} q(\hat{r}_{t+1} | \hat{a}_{t+1}, \hat{r}_{t+1}; \psi^2) q(\hat{a}_{t+1} | \hat{a}_{t+1}; \psi^3) q(\hat{a}_{t+1}; \psi^1) $$

(6.8)

These equations hold independently of which agent acts next in a specific interaction sequence, and we hence write $q(\hat{r}_{t+1} | \hat{r}_{t}; \phi^2)$ instead of $q(\hat{r}_{t+1} | \hat{r}_{t}; \phi_C)$ and $q(\hat{r}_{t+1} | \hat{r}_{t}; \psi^2)$ instead of $q(\hat{r}_{t+1} | \hat{r}_{t}; \phi_C)$.

The state-dependent player empowerment uses the NPC’s estimate of the player’s potential influence on their perceptions $n$ interaction cycles ahead. For this to be accurate, the NPC must use their beliefs about the player’s beliefs about their own and about the dynamics of the NPC and ‘other’ agent:

$$ \mathcal{E}^P(\hat{r}_{t}; \theta_P, \phi_P, \psi_P) = \max_{q(\hat{r}_{t+c})} I(\hat{A}_{t+c} \rightarrow S_{t+1+T}; \hat{r}_{t+c}, \theta_P, \phi_P, \psi_P) $$

(6.9)

Empowerment is calculated as the channel capacity of the $n$-step predictive factor with respect to the player’s future sensor states:

$$ q(S_{t+1+T} | \hat{a}_{t+1+T}, \hat{r}_{t+1+T}; \theta_P, \phi_P, \psi_P) = \sum_{\hat{r}_{t+1+T} \hat{a}_{t+1+T}} q(S_{t+1+T} | \hat{r}_{t+1+T}, \hat{a}_{t+1+T}; \phi_P) $$

$$ q(\hat{r}_{t+1+T} | \hat{a}_{t+1+T}, \hat{r}_{t+1+T}; \theta_P, \phi_P, \psi_P) $$

(6.10)

The $n$-step cyclic dynamics are given by Eq. 6.5, and use the following 1-step cyclic dynamics calculated from the player’s perspective:

$$ q(\hat{r}_{t+1} | \hat{a}_{t+1}, \hat{r}_{t+1}; \theta_P, \phi_P, \psi_P) = \sum_{\hat{r}_{t+1} \hat{a}_{t+1}} q(\hat{r}_{t+1} | \hat{a}_{t+1}, \hat{r}_{t}; \theta_P) q(\hat{a}_{t+1} | \hat{a}_{t+1}; \phi_P^2) $$

$$ q(\hat{a}_{t+1} | \hat{a}_{t+1}, \hat{r}_{t+1}; \phi_P^2) $$

(6.11)

The last term is the player’s model of their environment dynamics. The middle term represents the impact of the ‘other’ agent on the latent environment state (Eq. 6.8), and the first term accounts for the NPC’s impact, given by:

$$ q(\hat{r}_{t+1} | \hat{r}_{t}; \theta) = \sum_{\hat{a}_{t+1} \hat{r}_{t+1}} q(\hat{r}_{t+1} | \hat{a}_{t+1}, \hat{r}_{t+1}; \theta^2) q(\hat{a}_{t+1} | \hat{a}_{t+1}; \theta^3) q(\hat{a}_{t+1}; \theta^1) $$

(6.12)

Again, this distribution can be computed for the beliefs of different agents, and we thus omit agent-specific indices.
While the NPC and player empowerment can be considered straightforward extensions of vanilla empowerment, NPC-player transfer empowerment is special in that it only works in a multi-agent context. The state-dependent NPC-player transfer empowerment quantifies how much the NPC can potentially influence the player’s future perceptions with n-step interleaved action sequences. It relies on the NPC’s beliefs about their own and the dynamics of the player and the ‘other’ agent, and is given by:

\[
\mathcal{C}^\text{CP}(\hat{r}_t; \theta_C, \phi_C, \psi_C) = \max_{q(\hat{a}^\text{CP}_t)} \, I(\hat{A}^C_{t,n} \rightarrow \hat{S}_t^{p(n-1)} | \hat{r}_t; \theta_C, \phi_C, \psi_C) \quad (6.13)
\]

In contrast to the former variants, the mutual information here is calculated on n-step cyclic dynamics of the player and the ‘other’ agent, and is given by:

\[
q(\hat{s}_t^{p(n-1)} | \hat{a}^C_{t,n}, \hat{r}_t; \theta_C, \phi_C, \psi_C) = \sum_{l_{t,n}} q(\hat{s}_t^{p(n-1)} | \hat{r}_t^{p(n-1)}, \hat{r}_t^{1}) q(\hat{r}_t^{p(n-1)}, \hat{r}_t; \theta_C, \phi_C, \psi_C) \quad (6.14)
\]

Because the NPC precedes the player in the same interaction cycle, their action consequences must be assessed on the player’s future sensor n – 1, rather than n cycles ahead. Due to the NPC’s different position in the interaction cycle, the required n-step dynamics are not cyclic. For n > 1, we have:

\[
q(\hat{r}_t^{p(n-1)}, \hat{a}^C_{t,n}, \hat{r}_t; \theta_C, \phi_C, \psi_C) = \sum_{l_{t,n}} q(\hat{r}_t^{p(n-1)} | \hat{a}^C_{t,n}, \hat{r}_t^{c+1(n-1)}, \hat{r}_t^{c+1(n-1)}, \theta_C^2) q(\hat{r}_t^{c+1(n-1)} | \hat{a}^C_{t,n}, \hat{r}_t; \theta_C, \phi_C, \psi_C) \quad (6.15)
\]

The first term here represents the NPC’s environment dynamics model, and the second the n-step cyclic dynamics as recursively defined in Eq. 6.5. These in turn are based on the NPC’s 1-step cyclic dynamics given by Eq. 6.6.

The coupled empowerment action-value function forms the basis of the CEM policy. It associates all actions that the NPC could perform next with the coupled empowerment they are expected to yield. This is accomplished by combining the expected NPC, player and NPC-player transfer empowerment. These individual expectations not only depend on the potential future latent environment states that a specific action could cause, but also on the NPC’s beliefs and the latent environment states which they might presently occupy. We illustrate the calculation of the coupled empowerment action-value in Fig. 6.3. It shows the generative model unrolled over two interaction cycles, starting with the NPC at time \( t_C = t \). We distinguish four main stages.

The NPC firstly (i) has to account for the present latent environment state and beliefs in the expectation. To this end, they plug all sensorimotor experience up to time \( t_C \), i.e. \( m^C_{t_C} = (s_0, a_0, s_1, \ldots, a_{t_C-1}, s_{t_C}) \), into the generative model, and use it to infer the posterior factor (Eq. 6.1) over the latent environment states \( \hat{r}_{t_C} \) and parameters \( \theta, \phi, \psi \), based on the fixed hyperparameters \( \xi^*, \sigma^*, \omega^* \). Since the memorised sensor and action values have been directly
Figure 6.3: Calculation of the (non-simplified) 1-step coupled empowerment action-value in the generative model. Hyperparameters have been fixed to $\xi, \sigma, \omega$, and sensorimotor experience $a_{<t}, s_{<t}$ has been included up to time $t$. The coupled empowerment calculation relies on the NPC anticipating (--) the possible future latent environment states $\hat{R}_{t+1}, \hat{R}_{t+3}$ when the player (P) and NPC (C) act next. For each anticipated state $\hat{R}_{t+1}$, the 1-step player empowerment (--) is calculated by considering the impact of the player’s actions on their perception after one interaction cycle. Similarly, for each state $\hat{R}_{t+3}$, the NPC empowerment (--) and the NPC-player transfer empowerment (--) is calculated as the NPC’s potential influence on their own and the player’s future perceptions, respectively.

observed and performed, they are not hatted in Fig. 6.3. The NPC secondly (ii) anticipates which potential future latent environment states their assumed action $\hat{a}_C^t$ might produce when the player acts next at $t_P = t + 1$, and when they perform again themselves at $t_C + \tau = t + 3$. We illustrate these two anticipation steps in Fig. 6.3 with dashed arrows (--) from $\hat{a}_C^t$ to the latent environment states $\hat{R}_{t+1}$ and $\hat{R}_{t+3}$. In the first case, the NPC must only predict their own impact on the shared environment, but in the second case they must also account for the impact of the player and ‘other’ character interleaving their actions (Eq. 6.5). Thirdly (iii), the NPC calculates the $n$-step player empowerment (--) for all potential environment states $\hat{R}_{t+1}$ which they might take the player into. Moreover, they compute their own $n$-step NPC empowerment (--) and the NPC-player transfer empowerment (--) for each potential latent environment state $\hat{R}_{t+3}$ which they predict to be themselves in, after one interaction cycle. Due to space limitations, Fig. 6.3 only illustrates the calculation of 1-step empowerment. Finally (iv), the NPC calculates the expectation for each of these empowerment variants, and combine these individual expectations in the action-value for $\hat{a}_C^t$.
Since the expectations for the individual empowerment variants rely on inferring the posterior factor (Eq. 6.1), they all depend on the NPC’s memory and the fixed hyperparameters. The expected NPC empowerment is:

$$\mathbb{E}^C(\hat{a}_{tC}^C, m_{tC}^C, \xi_C, \sigma_C, \omega_C) = \mathbb{E}_{\hat{r}_{tC+\tau}, \hat{r}_{tC}, \Phi_C, \Psi_C, \Phi_C, \Psi_C} \mathbb{E}_{\hat{a}_{tC}^C, \hat{r}_{tC}, \hat{r}_{tC}, \Phi_C, \Psi_C} \left[ \mathbb{E}^C \right] \quad (6.16)$$

The last term in Eq. 6.17 is the NPC’s empowerment (Eq. 6.2) in the latent environment states when they act next, i.e. one interaction cycle ahead. The middle term is the posterior factor (Eq. 6.1), and the first term is the NPC’s 1-step cyclic dynamics given by Eq. 6.6.

The expected player empowerment is calculated over each potential present environment state and parameter, as well as the future latent environment state in which the player could act next, resulting from a specific NPC action:

$$\mathbb{E}^P(\hat{a}_{tC}^C, m_{tC}^C, \xi_C, \sigma_C, \omega_C) = \mathbb{E}_{\hat{r}_{tC}, \hat{r}_{tC}, \Phi_C, \Psi_C} \mathbb{E}_{\hat{a}_{tC}^C, \hat{r}_{tC}, \phi_C, \Psi_C} \left[ \mathbb{E}^P \right] \quad (6.18)$$

The player empowerment (Eq. 6.9) is calculated from the player’s perspective and thus relies on the player-associated set of parameters \( \theta_F, \phi_F, \psi_F \). Crucially though, the anticipation of the states \( r_{tF} \) in which the player acts next, induced by the NPC’s action, relies on the NPC’s environment dynamics and we thus also incorporate the corresponding beliefs \( \theta_C^2 \).

The expected NPC-player transfer empowerment is calculated in close analogy to the expected NPC empowerment: the expectation is taken over the same future environment states when the NPC can act next, but with respect to the NPC-player transfer empowerment (Eq. 6.13):

$$\mathbb{E}^{CP}(\hat{a}_{tC}^C, m_{tC}^C, \xi_C, \sigma_C, \omega_C) = \mathbb{E}_{\hat{r}_{tC+\tau}, \hat{r}_{tC}, \Phi_C, \Psi_C} \mathbb{E}_{\hat{a}_{tC}^C, \hat{r}_{tC}, \hat{r}_{tC}, \Phi_C, \Psi_C} \left[ \mathbb{E}^{CP} \right] \quad (6.20)$$
The coupled empowerment action-value function is given by a linear combination of all three empowerment types:

\[
E_{CEM}(\hat{a}_t^C, m_t^C; \alpha, \xi, \sigma, \omega) = \alpha_C E_C(\hat{a}_t^C, m_t^C; \xi_C, \sigma_C, \omega_C) + \alpha_P E_P(\hat{a}_t^C, m_t^C; \xi_P^C, \sigma_P, \omega_P) + \alpha_{CP} E_{CP}(\hat{a}_t^C, m_t^C; \xi_C, \sigma_C, \omega_C)
\]

Each empowerment variant is weighed by a coefficient \( \alpha \in [0, 1] \) as hyperparameters of the CEM model. They are summarised in the set \( \alpha = (\alpha_C, \alpha_P, \alpha_{CP}) \).

We finally formalise the subjective CEM policy as the agent-internal counterpart to the objective policy defined in the multi-agent PA-loop with memory (Fig. 6.1a). As for vanilla EM (Eq. 3.18), we formalise CEM based on greedy action selection with added stochasticity:

\[
q(\hat{a}_t^C | m_t^C) = \begin{cases} 
\frac{1}{|\hat{A}^C_t(m_t^C)|} & \text{if } \hat{a}_t^C \in \hat{A}^{C,*}(m_t^C), \\
0 & \text{otherwise.}
\end{cases}
\]

The set \( \hat{A}^{C,*}(m_t^C) \) here comprises all assumed actions that equally maximise coupled empowerment:

\[
\hat{A}^{C,*}(m_t^C) = \arg \max_{\hat{a}_t^C} E_{CEM}(\hat{a}_t^C, m_t^C; \alpha, \xi, \sigma, \omega)
\]

We next introduce several assumptions to simplify the calculation of coupled empowerment as an IR and CEM as a social IM model.

### Simplified Coupled Empowerment Maximisation

We want to provide a proof-of-concept of CEM’s potential for engineering more general and believable, supportive and adversarial NPCs. To this end, we formalise a simplified account of coupled empowerment by suspending some tough challenges that this social IM model poses, as they do not necessarily apply in videogames. As game engineers, we can obtain full access to the game state, the complete or a simplified forward model and the dynamics of all involved characters; we thus do not have to consider how the corresponding beliefs could be inferred from the NPC’s experience. We make the following assumptions to overcome the need for inference:

**Fixed parameters:** We assume that the beliefs over the initial environment state as well as all parameters for the NPC’s own, the player’s and the ‘other’ character’s sensor dynamics, environment dynamics and policy have been acquired beforehand and remain permanently fixed to \( \theta^* = (\theta^{0,*}, \theta^{1,*}, \theta^{2,*}, \theta^{3,*}) \) for the NPC, \( \phi^* = (\phi^{1,*}, \phi^{2,*}, \phi^{3,*}) \) for the player, and \( \psi^* = (\psi^{1,*}, \psi^{2,*}, \psi^{3,*}) \) for the ‘other’ character. The parameters are thus delta-distributed, i.e. \( q(\theta^i | \theta^i) = \delta_{\theta^i, \theta^{i,*}} \delta(\xi^i - \xi^{i,*}) \) for \( i \in [0, 3] \), and \( q(\phi^i | \phi^i) = \delta_{\phi^i, \phi^{i,*}} \delta(\omega^i - \omega^{i,*}) \) for \( i \in [1, 3] \).

**Known environment state:** We assume that the present latent environment state \( r_t \) is known to the NPC and used for action selection. They
must however still estimate the probability of future states \( \hat{r}_{t+1} \) based on their model of their own and the other characters’ dynamics.

The first assumption has been adopted from the formalisation of vanilla EM in Sec. 3.2. We add the second assumption to overcome the need for inference and yet maintain partial observability. The latter is crucial to keep the NPC and player empowerment separate and demonstrate the rewards’ sensitivity to the embodiment of each individual character.

Without inference, there is no need for memory and we thus represent the objective multi-agent interaction of the NPC, player and ‘other’ character by means of the memoryless PA-loop in Fig. 6.1b. The network topology also reflects our assumption that the NPC can directly access the latent environment state, and uses it to select their next action. We introduce two more assumptions to further reduce the model’s complexity:

**PERSPECTIVE COLLAPSE**: We assume that parametrisations of the same dynamics match across perspectives, i.e. \( \theta_{C}^{1,2} = \theta_{P}^{1,2}, \phi_{C}^{1,2} = \phi_{P}^{1,2}, \psi_{C}^{1,2,3} = \psi_{P}^{1,2,3} \). We omit the agent-specific indices for these parameters.

**PERFECT SENSORIMOTOR MODELS**: We assume that the sensor and environment dynamic models match the objective dynamics perfectly, e.g. \( q(\hat{s}_{C}^{C}|\hat{r}_{C}; \theta^{1}) = p(s_{C}^{C}|r_{C}) \) and \( q(\hat{r}_{C+1}|\hat{a}_{C}^{C}, \hat{r}_{C}; \theta^{2}) = p(\hat{r}_{C+1}|a_{C}^{C}, r_{C}) \). We omit these parameters from the corresponding model distributions.

In order to investigate the role of model accuracy for the emerging behavioural dynamics, and to assess the robustness of CEM with respect to sub-optimal models, we exclude the policies of NPC, player and the ‘other’ agent from these assumptions. We for instance do not assume the NPC to know the player’s true policy. We retain the four corresponding parameters \( \theta_{P}^{3}, \phi_{C}^{3}, \psi_{C,P}^{3} \) in the following formalisation, and investigate how changes to these parameters affect the emerging interaction in our later studies.

The impact of these assumptions on the CEM formalism is most visible in the expectations over the different empowerment variants for the action-value function. More subtly, the other equations simplify in that they depend on fewer loose parameters. The state-dependent NPC empowerment becomes:

\[
\mathcal{C}^{C}(\hat{r}_{C}; \phi_{C}^{3}, \psi_{C}^{3}) = \max_{\hat{S}_{C}^{C}} I(\hat{A}_{C}^{C,\pi} \rightarrow \hat{S}_{C}^{C+\pi}; \hat{r}_{C}; \phi_{C}^{3}, \psi_{C}^{3})
\]  

(6.26)

The simplified n-step predictive factor relies on two instead of seven parameters:

\[
q(\hat{s}_{C}^{C+\pi}|\hat{a}_{C}^{C,\pi}, \hat{r}_{C}; \phi_{C}^{3}, \psi_{C}^{3}) = \sum_{\hat{r}_{C+\pi}} q(\hat{s}_{C}^{C+\pi}|\hat{r}_{C+\pi}) q(\hat{r}_{C+\pi}|\hat{a}_{C}^{C,\pi}, \hat{r}_{C}; \phi_{C}^{3}, \psi_{C}^{3})
\]  

(6.27)

For \( n > 1 \), the simplified n-step cyclic dynamics are:

\[
q(\hat{r}_{t+\pi}|\hat{a}_{t}^{n-1}, \hat{r}_{t}; \ldots) = \sum_{\hat{r}_{t+(n-1)\pi}} q(\hat{r}_{t+\pi}|\hat{a}_{t+(n-1)\pi}, \hat{r}_{t+(n-1)\pi}; \ldots) q(\hat{r}_{t+(n-1)\pi}|\hat{a}_{t}^{n-1}, \hat{r}_{t}; \ldots)
\]  

(6.28)
The dots represent placeholders for parameter pairs $\theta^3, \psi^3$ and $\phi^3, \psi^3$, depending on the agent for which the cyclic dynamics are calculated. For the NPC, the $\tau$-step cyclic dynamics in Eq. 6.28 simplify to:

$$q(\hat{r}_{t+\tau} | \hat{a}_{t+\tau}^C, \hat{r}_{t+\tau}; \Phi^C, \Psi^C) = \sum_{\hat{r}_{t}, \hat{r}_{t+\tau}} q(\hat{r}_{t+\tau} | \hat{r}_t; \Psi^C) q(\hat{r}_t | \hat{a}_{t+\tau}^C, \hat{r}_{t+\tau})$$

(6.29)

Assuming that the parametrisation of environment and sensor dynamics is fixed and yields perfect models, the distributions modelling the player’s and ‘other’ agent’s impact on the latent environment state simplify to:

$$q(\hat{r}_{t+\tau}+1 | \hat{r}_{t+\tau}; \theta^3_P, \psi^3_P) = \sum_{\hat{a}_{t+\tau}^P, \hat{s}_{t+\tau}^P} q(\hat{r}_{t+\tau+1} | \hat{a}_{t+\tau}^P, \hat{r}_{t+\tau}) q(\hat{a}_{t+\tau}^P | \hat{s}_{t+\tau}^P; \theta^3_P, \psi^3_P) q(\hat{s}_{t+\tau}^P | \hat{r}_{t+\tau})$$

(6.30)

$$q(\hat{r}_{t+\tau}+1 | \hat{r}_{t+\tau}; \psi^3) = \sum_{\hat{a}_{t+\tau}^O, \hat{s}_{t+\tau}^O} q(\hat{r}_{t+\tau+1} | \hat{a}_{t+\tau}^O, \hat{r}_{t+\tau}) q(\hat{a}_{t+\tau}^O | \hat{s}_{t+\tau}^O; \psi^3) q(\hat{s}_{t+\tau}^O | \hat{r}_{t+\tau})$$

(6.31)

In a similar manner, the NPC’s calculation of the state-dependent player empowerment simplifies to:

$$e^P (\hat{r}_{t+\tau}; \theta^3_P, \psi^3_P) = \max_{q(\hat{a}_{t+\tau}^P)} I(\hat{A}_{t+\tau}^P \rightarrow \hat{S}_{t+\tau+nT}^P | \hat{r}_{t+\tau}; \theta^3_P, \psi^3_P)$$

(6.32)

It is based on the simplified $n$-step predictive factor:

$$q(\hat{s}_{t+nT} | \hat{a}_{t}^{P,n}, \hat{r}_{t+\tau}; \theta^3_P, \psi^3_P) = \sum_{\hat{r}_{t+nT}} q(\hat{s}_{t+nT} | \hat{r}_{t+nT}) q(\hat{r}_{t+nT} | \hat{a}_{t}^{P,n}, \hat{r}_{t+\tau}; \theta^3_P, \psi^3_P)$$

(6.33)

The $n$-step cyclic dynamics are given by Eq. 6.28, using the $\tau$-step cyclic dynamics:

$$q(\hat{r}_{t+\tau} | \hat{a}_{t}^P, \hat{r}_{t+\tau}; \theta^3_P, \psi^3_P) = \sum_{\hat{r}_{t+\tau}} q(\hat{r}_{t+\tau} | \hat{a}_{t}^P, \hat{r}_{t+\tau}; \theta^3_P, \psi^3_P) q(\hat{r}_{t+\tau} | \hat{a}_{t}^P, \hat{r}_{t+\tau})$$

(6.34)

The impact of the ‘other’ agent on the latent environment state is calculated from Eq. 6.31. The distribution encoding the NPC’s impact simplifies to:

$$q(\hat{r}_{t+1} | \hat{r}_t; \theta^3) = \sum_{\hat{a}_{t}^C, \hat{s}_{t}^C} q(\hat{r}_{t+1} | \hat{a}_{t}^C, \hat{r}_t) q(\hat{a}_{t}^C | \hat{s}_{t}^C; \theta^3) q(\hat{s}_{t}^C | \hat{r}_t)$$

(6.35)

In contrast to the non-simplified version, this only relies on a single parameter: the player’s beliefs about the NPC’s policy.

The simplified state-dependent NPC-player transfer empowerment is:

$$e^{CP} (\hat{r}_{t}; \Phi^C, \Psi^C) = \max_{q(\hat{a}_{t}^C)} I(\hat{A}_{t}^C \rightarrow \hat{S}_{t+1}^P | \hat{r}_t; \Phi^C, \Psi^C)$$

(6.36)
The n-step predictive factor, mapping from the NPC’s n-step action sequences to the player’s future perceptions, becomes:

\[
q(s_{tp+(n-1)\tau}|a_{t_c}^{C,n}, \hat{r}_{t_c}; \phi_C^3, \psi_C^3) = \sum_{r_{tp+(n-1)\tau}} q(s_{tp+(n-1)\tau}|r_{tp+(n-1)\tau})
\]

\[
q(\hat{r}_{tp+(n-1)\tau}|a_{t_c}^{C,n}, \hat{r}_{t_c}; \phi_C^3, \psi_C^3)
\] (6.37)

The last term, i.e. the acyclic n-step dynamics simplify to:

\[
q(\hat{r}_{tp+(n-1)\tau}|a_{t_c}^{C,n}, \hat{r}_{t_c}; \phi_C^3, \psi_C^3) = \sum_{r_{tp+(n-1)\tau}} q(\hat{r}_{tp+(n-1)\tau}|a_{t_c}^{C,n}, r_{t_c}; \phi_C^3, \psi_C^3) \psi_C^3
\]

\[
q(\hat{r}_{t_c+(n-1)\tau}|a_{t_c}^{C,n}, \hat{r}_{t_c}; \phi_C^3, \psi_C^3)
\] (6.38)

The first term here represents the NPC’s exact environment dynamics model, and the second the simplified n-step cyclic dynamics in Eq. 6.28.

Assuming knowledge of the present latent environment state \(r_{t_c}\) and fixed environment and sensor dynamics parameters, the complexity of the expectations required for the coupled empowerment action-value function is considerably reduced. In the simplified version, they are only calculated over the potential future latent environment states caused by the NPC’s actions. The expected NPC, player and NPC-player transfer empowerment becomes:

\[
\mathbb{E}^C(\hat{a}_{t_c}^{C}, r_{t_c}; \theta_C^3, \psi_C^3) = \mathbb{E}_{R_{t_c}}[\psi_{t_c}^{C}\phi_{t_c}^{3}e_C]\] (6.39)

\[
\mathbb{E}^P(\hat{a}_{t_c}^{C}, r_{t_c}; \theta_P^3, \psi_P^3) = \mathbb{E}_{R_{t_c}}[\psi_{t_c}^{P}e_P]\] (6.40)

\[
\mathbb{E}^{CP}(\hat{a}_{t_c}^{C}, r_{t_c}; \theta_P^3, \psi_P^3) = \mathbb{E}_{R_{t_c}}[\psi_{t_c}^{P}e_{CP}]\] (6.41)

\[
\mathbb{E}^{C EM}(\hat{a}_{t_c}^{C}, r_{t_c}; \alpha, \theta_P^3, \theta_C^3, \psi_P^3, \psi_C^3) = \alpha_C \mathbb{E}^C(\hat{a}_{t_c}^{C}, r_{t_c}; \theta_C^3, \psi_C^3)
\]

\[
+ \alpha_P \mathbb{E}^P(\hat{a}_{t_c}^{C}, r_{t_c}; \theta_P^3, \psi_P^3)
\]

\[
+ \alpha_{CP} \mathbb{E}^{CP}(\hat{a}_{t_c}^{C}, r_{t_c}; \theta_P^3, \psi_P^3)\] (6.42)
Algorithm 1: Coupled Empowerment Maximisation (CEM)

Stage 1: Anticipation
1: function \( \pi(r_{iC}, \alpha, \theta^3_P, \phi^3_C, \psi^3_P) \)
2: for all \( \hat{a}^C_{iC} \in \hat{A}^C_{iC} \) do
3: \( q(\hat{r}_{tp} | a^C_{iC}, r_{iC}) \)
4: \( q(\hat{r}_{iC+\tau} | a^C_{iC}, r_{iC}; \phi^3_C, \psi^3_C) \) (Eq. 6.29), using \( q(\hat{r}_{tp} | a^C_{iC}, r_{iC}) \)
5: end for

Stage 2: Player Empowerment Calculation
6: for all \( \hat{r}_{tp} \in \hat{R}_{tp} : q(\hat{r}_{tp} | a^C_{iC}, r_{iC}) > 0 \) and \( a^{P,n}_{tp} \in \hat{A}^P_{iC} \) do
7: Calculate n-step dynamics \( q(\hat{r}_{tp+nT} | a^{P,n}_{tp}, \hat{r}_{tp}; \theta^3_P, \psi^3_P) \) (Eq. 6.28)
8: Calculate predictive factor \( q(\hat{s}^P_{tp+nT} | a^{P,n}_{tp}, \hat{r}_{tp}; \theta^3_P, \psi^3_P) \) (Eq. 6.33)
9: end for

Stage 2: NPC and NPC-Player Transfer Empowerment Calculation
10: for all \( \hat{r}_{iC+\tau} \in \hat{R}_{iC+\tau} : q(\hat{r}_{iC+\tau} | a^C_{iC}, \hat{r}_{iC}; \phi^3_C, \psi^3_C) > 0 \) and \( a^{C,n}_{iC+\tau} \in \hat{A}^C_{iC+\tau} \) do
11: Calculate recursively \( q(\hat{r}_{tp+nT} | a^{C,n}_{iC+\tau}, \hat{r}_{iC+\tau}; \phi^3_C, \psi^3_C) \) (Eq. 6.38)
12: Calculate predictive factor \( q(\hat{s}^P_{tp+nT} | a^{C,n}_{iC+\tau}, \hat{r}_{iC+\tau}; \phi^3_C, \psi^3_C) \), the NPC-player transfer predictive factor (Eq. 6.37), using \( q(\hat{r}_{tp+nT} | a^{C,n}_{iC+\tau}, \hat{r}_{iC+\tau}; \phi^3_C, \psi^3_C) \)
13: Calculate \( q(\hat{s}^C_{iC+(n+1)T} | a^{C,n}_{iC+\tau}, \hat{r}_{iC+\tau}; \phi^3_C, \psi^3_C) \), the NPC predictive factor (Eq. 6.27), using \( q(\hat{r}_{iC+(n+1)T} | a^{C,n}_{iC+\tau}, \hat{r}_{iC+\tau}; \phi^3_C, \psi^3_C) \)
14: end for

Stage 3: Coupled Empowerment Action-Value Calculation
15: Find \( q^e_C(a^{C,n}_{iC+\tau}) \) and \( q^e_C(a^{C,n}_{iC+\tau}) \) that maximise the channel capacity for both predictive factors with the Blahut-Arimoto algorithm
16: Calculate and store \( \mathcal{E}^C_P(\hat{r}_{iC+\tau}; \phi^3_C, \psi^3_C) \) (Eq. 6.36) using \( q^e_C(a^{C,n}_{iC+\tau}) \), and \( \mathcal{E}^C(\hat{r}_{iC+\tau}; \phi^3_C, \psi^3_C) \) (Eq. 6.26), using \( q^e_C(a^{C,n}_{iC+\tau}) \)
17: end for

Stage 4: Action Selection
18: Calculate \( q(\hat{a}^C_{iC} | r_{iC}) \), the greedy CEM policy (Eq. 6.46)
19: \( \text{Return } a^*_{iC} \sim q(\hat{a}^C_{iC} | r_{iC}) \)
20: end function
We finally formalise the simplified subjective CEM policy. Assuming the present environment state to be known, the policy is conditioned on \( r_{t_c} \) rather than the current sensor state \( s_{t_c} \):

\[
q(\hat{a}_{t_c}^C| r_{t_c}) = \begin{cases} 
\frac{1}{|\hat{A}_{t_c}^C(r_{t_c})|} & \text{if } \hat{a}_{t_c}^C \in \hat{A}_{t_c}^{C,*}(r_{t_c}), \\
0 & \text{otherwise.} 
\end{cases}
\] (6.46)

The set \( \hat{A}_{t_c}^{C,*}(r_{t_c}) \) of coupled empowerment maximising actions is given by:

\[
\hat{A}_{t_c}^{*}(r_{t_c}) = \arg \max_{\hat{a}_{t_c}^C} e_{C}^{EM}(\hat{a}_{t_c}^C, r_{t_c}; \theta_p^3, \psi_C^3, \psi_{P,c}^3)
\] (6.47)

We illustrate the computation of the simplified CEM policy with the pseudocode in Alg. 1. It draws on the Blahut-Arimoto algorithm (Arimoto, 1972; Blahut, 1972) for the calculation of the channel capacity. By following the same structure, our pseudocode complements the detailed description of the generic coupled empowerment action-value calculation in the previous section. It moreover allows us to emphasise how terms can be re-used, in particular in the calculation of NPC and NPC-player transfer empowerment. The cyclic and acyclic n-step dynamics (line 12) underlying these empowerment types share the same terms, and can be iteratively constructed without redundancy. The calculation of the n-step dynamics for the player empowerment however cannot benefit from this, as it relies on a different set of parameters. We use this algorithm, realising the simplified CEM policy, throughout our studies in the next section.

We briefly relate CEM to the work of Jaques et al. (2019), which has been out of scope for our systematic review of IM in game AI. Inspired by the transfer empowerment introduced in this section, Jaques et al. (ibid.) propose an intrinsic ‘social influence’ reward to increase coordination and communication in sequential social dilemma games in which agents must cooperate to achieve maximum overall score. In contrast to transfer empowerment, their reward is given by the mutual information between the actions of two agents, rather than the actions of one agent and the other’s sensory futures. While we rely on an exhaustive computation of rewards, they perform an RL-based Monte-Carlo approximation. They do not optimise an agent’s potential influence on the other’s sensor, but maximise their actual influence on the other’s actuator. Furthermore, they investigate bi-directional co-operation instead of uni-directional support. Most crucially, the co-operative behaviour of their agents is ultimately based on maximising extrinsic reward. CEM, in contrast, relies exclusively on IRs.

6.5 STUDIES: COUPLED EMPOWERMENT MAXIMISATION FOR GENERAL, BELIEVABLE COMPANION AND ADVERSARY CHARACTERS

We have conducted two observational vignette studies to assess the capacity of CEM to drive the behaviour of general, believable NPCs that either support the player as companions, or challenge them as adversaries. We consider general
NPCs a special case of co-creative agents (cf. Sec. 6.2), and CEM an instance of social models of IM. Our study seeks to contribute to the overarching research question of this chapter: ‘Can we use a model of intrinsic motivation to engineer general and social co-creative agents?’ (RQ.8).

To establish a proof-of-concept, we probe five predictions\(^ {16}\) on the behaviour of CEM-driven NPCs. The first two concern the emerging social dynamics:

- **PD.1** Maximising the player’s empowerment through CEM yields supportive NPC behaviour.
- **PD.2** Minimising the player’s empowerment through CEM yields antagonistic NPC behaviour.

The other three predictions relate to the NPC’s generality:

- **PD.3** A CEM-driven NPC responds sensibly to the behaviours of the player and other characters.
- **PD.4** A CEM-driven NPC responds with new but sensible behaviour to changes in their embodiment.
- **PD.5** A CEM-driven NPC responds with new but sensible behaviour to changes in their environment.

By ‘sensible behaviour’ we denote support or antagonism towards the player that is consistent with the targeted social dynamics.

We introduce the concept of observational vignettes as a qualitative method for the study of AI, inspired by experiential vignettes (Hudson & Cairns, 2014a) in games user research. Experiential vignettes are small-scale qualitative studies capable of shedding light on a little explored user experience phenomenon by evaluating participants’ responses to explicitly manipulated conditions. Observational vignettes follow the same idea, but are set-up to gather qualitative data in the form of observations of simulated AI behaviour as the phenomenon of interest. We use them to investigate our predictions of the generality and social dynamics exhibited by CEM-driven NPCs.

Our studies comprise several vignettes, each designed to probe a specific aspect of our predictions. Each vignette in turn consist of several conditions as manipulations of a simple game and the CEM hyperparameters. The human experimenter controls the player avatar, and observes the behaviour emerging from the interaction with the CEM-driven NPC. Our studies are not only systematic but also exploratory, in that we explore the possible behavioural outcomes of these manipulations beyond our predictions.

We have chosen this qualitative approach because quantitative means to evaluate gameplay and player experience cannot yet capture NPC-induced support and antagonism reliably and in full detail. A quantitative assessment of the NPC’s impact on objective performance indices such as the game score would likely allow us to distinguish supportive from antagonistic behaviour, but not reveal its quality as perceived by the player.

\(^ {16}\) We probe predictions rather than hypotheses, as our study results are qualitative, and hypotheses are commonly related to quantitative evaluation.
We provide more detail on our method in the following sections, as there exist slight variations between the studies. The two studies are mainly distinguished by the type of social dynamics investigated: we probe the ability of CEM to yield supportive NPC behaviour (PD.1) in the first, and its potential to give rise to adversarial behaviour (PD.2) in the second. The studies also focus on different aspects of generality. In the first, we probe whether CEM-driven NPCs can respond in a general and yet sensible way to unknown player behaviour and other dynamic elements in the game world (PD.3). To this end, we introduce enemies to the NPC player interaction that provide opportunities for, but also challenge the NPC’s support. In the second study, we focus more on the generality of CEM-driven NPCs in regards to dealing with changes in their environment (PD.5) and embodiment (PD.4). We facilitate this via an extension of the simulation testbed shared between both studies.

6.5.1 Study 1: Companion Non-Player Characters

In our first study, we evaluate whether CEM can give rise to believable NPCs that consistently support (PD.1) the player as companions, while remaining general with respect to different player and enemy behaviours influencing the shared game world (PD.3). Before reporting our method and results, we define what we mean by companions, what makes them believable, and what makes their design challenging. This complements our general account of NPCs in Sec. 6.2, and informs the conditions for our study.

6.5.1.1 Believable Companion Non-Player Characters

An NPC can ‘provide the player with help in form of advice, directions, or resources as a sidekick, or it may fight alongside the player as an ally, or do both’ (Emmerich, Ring & Masuch, 2018, p. 142, emphasis added). While their role thus overlaps with that of other friendly NPC types, companions specifically are characterised by persistently accompanying the player throughout large parts of a game (Warpefelt & Verhagen, 2017).

Due to this persistence, companions contribute strongly to player experience (PX). They can serve as tutors, introduce the player into the game world and, if perceived as believable, can evoke emotional and social responses similar to human co-players (Emmerich, Ring & Masuch, 2018). They can support the suspension of disbelief and increase perceived realism and immersion, thus affecting replayability (Bailey & Katchabaw, 2008). Based on an online survey (N = 237), Emmerich, Ring and Masuch confirm that an NPC can make a game more interesting (M = 3.34, SD = 0.76). Moreover, players wish for more games to feature compelling companions (M = 3.23, SD = 0.83), and deem a well-made NPC a reason to play a game (M = 3.17, SD = 1.03). Unsurprisingly, companions such as Dogmeat from Fallout (Interplay Pro-
ductions, 1997) and Ellie from The Last of Us (Naughty Dog, 2013) became engraved into players’ collective memory.

The same persistence poses a considerable design challenge: ‘As companions accompany the player the whole time in many different situations and settings of the game, they have to be able to adapt to the changing context to maintain believability’ (Emmerich, Ring & Masuch, 2018, p. 143). If companions do not meet players’ expectations towards their believability, e.g. by becoming unsupportive and hence breaking characterhood, they can become a great source of annoyance and shatter our game immersion (Cerny, 2015). In their study, Emmerich, Ring and Masuch (2018) find that an annoying NPC can make players quit a game ($M = 2.43, SD = 1.21$). As an example, they refer to the dog Meeko in The Elder Scrolls V: Skyrim (Bethesda Game Studios, 2011) which has been frequently criticised for setting off traps, hence endangering the player, or getting killed by prematurely attacking enemies. Since companions must maintain their believability throughout the game, their integration through traditional game AI methods (cf. Sec. 6.2) is challenging and costly. We address this challenge through CEM.

Players have specific expectation towards the behaviour of a believable companion NPC. In their study, Emmerich, Ring and Masuch (2018) find that players on the one hand want companions to support them in reaching their goal in the game ($M = 3.24, SD = 0.92$). On the other hand though, players also want companions to act independently and on their own ($M = 3.13, SD = 0.91$), to follow their own objectives and goals ($M = 3.25, SD = 0.86$), and to contradict them by having ‘their own head’ (ibid., p. 149) ($M = 3.08, SD = 1.08$). Finally, players expect a believable NPC to react appropriately to the current game situation and their actions ($M = 3.64, SD = 0.63$), and to often interact with the game world and their character ($M = 3.51, SD = 0.74$). Based on these observations, we propose a minimal conceptualisation of believable companion behaviour in terms of three duties. To realise companion characterhood, the NPC must:

1. **Maintain Operational Proximity**: Act towards states where they can support the player in the future.

2. **Ensure Player Integrity**: Ensure that the player can continue pursuing their goals, and act against any force that would constrain these abilities.

3. **Ensure Their Own Integrity**: Secure their own existence and hence the ability to support the player in the long term.

The second (2) duty captures the player’s demand for support, (3) emphasises the NPC’s autonomy and own agenda, and (1) ensures that the NPC keeps supporting the player. We did not define any duties relative to a specific goal, as this would constrain the NPC’s generality. Being intrinsically motivated, their goal-directedness arises from their specific embodiment, based on the game world and mechanics, and from their interaction with the player.

These duties also comprise the need for believable NPCs to follow their own agenda. They are complemented by the other criteria for NPC believability presented in Sec. 6.2, namely to exhibit behavioural diversity and to be sensitive to their surroundings and body.
6.5 STUDIES: CEM FOR GENERAL, BELIEVABLE & SOCIAL NPCs

6.5.1.2 Method

To investigate our predictions PD.1 and PD.3, we conduct an exploratory study based on three observational vignettes, each probing one companion duty as operationalisation of believable support (PD.1). As conditions, we manipulate the level design of a custom game introduced below and the abilities of the game characters across five experiments across the vignettes. For each condition, we describe the emergent behaviour, highlight the contributions of the individual types of empowerment in the NPC’s policy, and demonstrate how they blend together. The player avatar is controlled by the experimenters. Since the player and enemy behaviour changes throughout the experiment, we implicitly assess player and NPC generality (PD.3).

Our model, introduced below, is essentially the CEM formalism from Sec. 6.4 integrated in our testbed. In the spirit of an exploratory study, we observe shortcomings of this vanilla version with respect to realising companion-like behaviour. We introduce modifications to overcome these shortcomings and yet maintain the generality of the model on the fly.

6.5.1.3 Testbed

Our study is set in a dedicated, minimal dungeon-crawler game in which the player, supported by an NPC, has to defeat enemies in a maze consisting of rooms connected by corridors. Agents interact in turn-wise order, starting with the companion, and followed by the player and an arbitrary number of enemies. All characters have health points, and can either move one step in each direction, shoot over a range of four tiles, or idle. They can only hit other characters within a certain range in their view direction, which changes with movement. The enemies act deterministically, in that they always shoot at or chase whichever character is closest, either the NPC or the player, but never other enemies. We provide an overview and brief justification of the level elements and character abilities in Tbl. 6.2 and 6.3, respectively.

We chose this game type for various reasons. Dungeon crawlers are traditionally discrete in time and space, which affords direct application of CEM as defined in Sec. 6.4 and simplifies the analysis of the emerging behaviour. Moreover, the core mechanics reflect the struggle for survival in nature. Given its biological motivation (cf. Sec. 3.1), empowerment as an IR underlying CEM is thus likely to align with the game’s goals. Dungeon crawlers traditionally rely on PCG and elements of chance, and therefore afford interesting challenges to a general NPC, to be addressed in future work. Classic examples such as Nethack (The NetHack DevTeam, 1987) and recent variants such as

<table>
<thead>
<tr>
<th>Sprite</th>
<th>Type</th>
<th>Description</th>
<th>Reason for inclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Floor</td>
<td>Foundation of level; accommodates avatar.</td>
<td>To indicate area where characters can freely move.</td>
</tr>
<tr>
<td></td>
<td>Wall</td>
<td>Immovable obstacle. Cannot be penetrated by attacks or other characters’ sight.</td>
<td>To structure level and provide choke points for specific interactions. Affords hiding.</td>
</tr>
</tbody>
</table>

Table 6.2: Dungeon-crawler level elements in our study of companion NPCs.
<table>
<thead>
<tr>
<th>Ability</th>
<th>Description</th>
<th>Reason for inclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Idle</td>
<td>Causes no change to the current game state.</td>
<td>Can serve as a fallback if other actions are disadvantageous.</td>
</tr>
<tr>
<td>Move</td>
<td>Move character into adjacent cell if there is no obstacle. Otherwise only changes orientation.</td>
<td>Allows exploration, hiding, and change of position as a reaction to other characters.</td>
</tr>
<tr>
<td>Range attack</td>
<td>Reduces health of first character in current direction and attack range. Damage and range are predefined.</td>
<td>Allows a character to attack other characters, both friend and foe.</td>
</tr>
</tbody>
</table>

Table 6.3: Character abilities in our study of companion NPCs.

*PixelDungeon* (Retronic Games, 2015) illustrate how our minimalistic testbed could be extended to further probe **CEM**’s generality.

### 6.5.1.4 Model

The **NPC** policy is calculated as in Alg. 1, based on the simplified coupled empowerment without inference defined in Sec. 6.4.3.3. An arbitrary number of deterministic enemies take the role of the ‘other’ character in the formalism. We have designed the **NPC** and player sensors $S_C, S_P$ to be local, non-overlapping and asymmetric to keep player and **NPC** empowerment sensitive to the individual character embodiment only and hence separate. We model locality by only accounting for level elements in a maximum distance of two units around a character. Other characters are represented in the sensor by an ID and their relative position. Sensors are non-overlapping, i.e. they only comprise a character’s own absolute position, rotation, and health. They are asymmetric, in that the player sensor also comprises the game status (running, lost, won). By default, all characters have two out of two health points.

We assume the sensor and environment dynamics to be deterministic. Adopting our simplifying assumptions from Sec. 6.4.3.3, the characters’ individual dynamics models match their peers’ and the objective dynamics. Only the four policy models, parametrised by $\theta_P^3, \theta_C^3, \psi_P^3, \psi_C^3$, are excluded from this. We assume throughout our study that both **NPC** and player model the enemies’ deterministic policy accurately. For now, we also make the default assumption that the **NPC** believes the players’ actions to be uniformly distributed and vice versa, i.e. $q(\hat{a}_P^P|s_P^P, \phi_C^3) \sim \mathcal{U}(|\hat{A}_P^P|)$ and $q(\hat{a}_C^C|s_C^C, \theta_P^3) \sim \mathcal{U}(|\hat{A}_C^C|)$. We furthermore set the lookahead to a default value of $n = 2$.

The only variables left to specify are the hyperparameters $\alpha$ that determine how much player, **NPC** and **NPC-player** transfer empowerment are weighed into the coupled empowerment reward. Selecting these parameters is not straight-forward, as companion characters specifically must realise a delicate balance between being supportive and yet independent from the player. The non-overlapping sensor would make our **NPC** strictly egocentric, if they only maximised their own empowerment. However, there is agreement that an **NPC** must not be too independent or egocentric: in a qualitative study conducted by Cerny (2015), a player said ‘I dislike that [the companion] prioritises getting to the exit herself over helping [me] first’ (ibid., p. 6). But
6.5 Studies: CEM for General, Believable & Social NPCs

Figure 6.4: Experiment 1. Adding NPC-player transfer empowerment to the IR allows the NPC to maintain operational proximity to the player, and hence to follow them through the narrow corridor.

an NPC must also not be too supportive of the player and care too little for themselves: Emmerich, Ring and Masuch (2018) find in their survey that ‘players neither seek for inferior companions that have to be protected, nor do they show a strong favor for mentoring companions. The highest agreement is found for companions who are coequal and able to take care of themselves’ (ibid., p. 148). We achieve this trade-off in our simulations by weighing the player’s empowerment the most by $a^P = 0.5$, the NPC’s own empowerment...
by $a^C = 0.2$, and the NPC-player transfer empowerment by $a^{CP} = 0.3$. We found these hyperparameters experimentally, and assume them by default in the following vignettes.

6.5.1.5 Duty 1: Maintain Operational Proximity

To support the player in achieving their goals, the NPC should strive for situations in which they can affect the player and their future perceptions best. Depending on the NPC’s action set, this operational proximity can be different from spatial proximity: We could image an NPC that can push buttons on a terminal, but cannot engage in close combat. Such an NPC might support the player most by staying remote, where it could e.g. unlock doors or trap enemies. Probing such operational proximity is subject to our first experiment. As condition, we have designed a level with two rooms connected by a narrow corridor. Given the characters’ default abilities and level design, operational proximity here comes down to spatial proximity, and we expect the NPC to stay close to the player and follow them from one room to the other.

However, such behaviour is not self-evident, as illustrated in Fig. 6.4. Here, the NPC and player are represented by violet and purple squares with letters ‘C’ and ‘P’, respectively. The numbers on the bottom specify their current and maximum health. The individual figures illustrate the empowerment rewards relevant to the policy, by mapping them as grey-scale values to different positions in the scene. Brighter hues indicate higher empowerment. They are calculated by fixing the player’s position and moving the north-facing NPC around. Hence, only the player avatar is shown. The value at a location in Fig. 6.4a and 6.4b corresponds to the NPC empowerment and NPC-player transfer empowerment, respectively, if the NPC was in that position.

Via this procedure, we find that the NPC’s empowerment (Fig. 6.4a) is particularly low at the room edges and corners, but also in the corridor. In these positions, the NPC can move neither north nor south and their sequences of navigational actions collapse into very few follow-up states. The corridor thus represents an NPC empowerment bottleneck. If the NPC’s policy was only about maximising their own empowerment, they would thus move to the centre of the current room, and avoid the corridors. NPC-player transfer empowerment (Fig. 6.4b), in contrast, renders all positions but the ones in which the NPC can influence the player’s perception directly as less attractive. When coupled with the other empowerment types (Fig. 6.4c), it compensates for the bottleneck induced by the NPC’s empowerment. For the default setup and $a_{CP} \geq 0.3$, the NPC consequently follows the player through the corridor and maintains spatial proximity, as shown in Fig. 6.4c.

The digital appendix contains a video documenting this essential behaviour.

Experiment 2: Trust

As a result of exploring the emerging behaviour in this experiment, we have identified a shortcoming of the vanilla CEM formalism with respect to driving companion NPCs: in specific situations, the player may be perceived as a threat by the NPC, which can disturb their operational proximity specifically and supportive behaviour more generally. We dedicate our second experiment to understanding and overcoming this shortcoming.

We reproduce a situation in which the player is not trusted in the simplest conceivable environment: a single room (Fig. 6.5a). We calculate the NPC’s
empowerment in this room for different positions of a north-facing NPC and a fixed player. The resulting empowerment landscape in Fig. 6.5 highlights that the NPC’s empowerment is particularly low when facing the player. This is due to the NPC modelling the player’s policy as uniform; in the NPC empowerment calculation (Eq. 6.26), they hence assume that the player would perform an action that reduced their sensory futures and hence empowerment with the same probability as any other action. The NPC consequently flees from the player, and avoids staying within their shooting range. We suggest that this assumption and the resulting behaviour are unnatural for the interaction of a supporting character and a player who would benefit from ongoing support. For supportive behaviour to emerge, it is important that player and NPC realise trust: they must not assume their peer to perform future actions that would significantly threaten their own existence.

We introduce a trust correction extension to overcome this shortcoming. It modifies the NPC’s model of the player’s policy such that actions that counteract trust are assigned zero probability and hence are not considered in the 1-step cyclic dynamics (Eq. 6.29) for the NPC empowerment calculation. This is accomplished in two steps. Firstly, a subset of trust-maintaining player actions \( \hat{A}_P \subseteq \hat{A} \) in the assumed environment state \( \hat{r}_p \) is identified as follows:

\[
\hat{A}_{P,x}^P(\hat{r}_p) := \left\{ \hat{a}_p^P \in \hat{A}_P^P : \mathcal{C}(\hat{r}_p; \hat{a}_P^P ; \psi_P^0) > 0 \land \sum_{\hat{r}_p' \in \mathcal{R}_P(\hat{r}_p, \hat{a}_P^P)} \mathcal{C}(\hat{r}_p' ; \phi_P^0, \psi_P^0) > 0 \right\} \tag{6.48}
\]

with

\[
\mathcal{R}_P(\hat{r}_p, \hat{a}_P^P) := \{ \hat{r}_{p+1} \in \mathcal{R}_{P+1} : q(\hat{r}_{p+1} | \hat{r}_p, \hat{a}_P) > \eta \}
\]
The set only retains player actions that do not directly reduce the NPC’s empowerment to zero with probability $\eta > 0$. To this end, we calculate the expected NPC empowerment over latent environment states $\hat{r}_t$ after the player has acted, and before the enemies can intervene. By comparing this to the NPC’s empowerment immediately before the player has performed in $\hat{r}_t$, we can infer whether it was the player’s action that has rendered NPC empowerment zero, rather than another character’s. The set still comprises player actions that reduce the NPC’s empowerment but are not fatal, as such actions might benefit the player in some other way. As a second step, this set is used to update the NPC’s model of the player’s policy such that trust-maintaining actions are assigned equal, and other actions zero probability:

$$q(\hat{a}^p_t, \hat{s}^p_t; \phi^3_C) = \begin{cases} \frac{1}{|\hat{A}_t^p|} & \text{if } \hat{a}^p_t \in \hat{A}_t^p, r_t, \\
0 & \text{otherwise.} \end{cases}$$

Crucially, this extension does not constrain the generality of the overall model, as it again relies on empowerment as IR. Applied to the previous scenario, the NPC empowerment remains high in front of the player and within their shooting range (Fig. 6.5a). Moreover, the CEM-driven NPC does not flee any more from the player. We have documented the initial fleeing behaviour in a video comprised in the digital appendix.

We conclude that a CEM-driven NPC can realise operational proximity supported by trust. In our testbed, this is articulated in the NPC following and remaining close to the player.

### 6.5.1.6 Duty 2: Ensure Player Integrity

In our testbed, ensuring the player’s integrity entails protecting them, and to prevent their death. Probing such behaviour is subject to our third experiment. For this purpose, we design a level scenario in which the player is directly threatened by an enemy, as illustrated in Fig. 6.6. The enemy, represented by an orange square and the letter ‘E’, faces the player and is ready to shoot. We illustrate the different empowerment types relevant to the policy in Fig. 6.6a and 6.6b. The value at a particular location in Fig. 6.6a corresponds to the player’s empowerment, if the north-facing NPC was in that position and chose to shoot. Since player empowerment quantifies their potential influence on future perceptions, it would drop to zero if the player was killed, as their sensory futures would collapse into one state. The player’s empowerment is thus highest if the NPC either faces the enemy in shooting range, or positions themselves between the two, to take the bullet. Fig. 6.6b reveals the latter alternative in a different way. It shows the NPC-player transfer empowerment, i.e. the NPC’s influence on the player’s future sensor state, for different NPC positions. As such, it highlights where the NPC could act as a bodyguard and save the player by stepping between them and the enemy from the side.

Maximising coupled empowerment leads to the NPC killing the enemy for any value of $\hat{a}_C$ and $\hat{a}_{CP}$, as long as $\hat{a}_P > 0$. If $\hat{a}_P = 0$, the NPC does not value the player’s empowerment at all. As their sensors do not overlap, the
NPC would hence not ‘care’ about the player. Importantly, the NPC would defend the player even if the enemies did not pose a threat to themselves.

Our exploration of this specific scenario shows that the NPC protects the player as long as they can anticipate the player’s death using their $n$-step lookahead. Even for a small lookahead $n = 1$, the NPC would always protect a player that only had one health point left. Crucially though, the NPC would stop protecting the player if the player’s present health exceeded their lookahead. The reason for this is an inconsistency in many (video-)games: In nature, a living being’s lower health not only indicates its closeness to death, but also corresponds to a decline in their ability to interact successfully with the world. In games though, health or similar labels for fitness often only represent a mere warning, and affect the character’s performance irregularly or only when dropping to zero. An NPC can thus only foresee the tragic consequences of the enemy’s actions if they evaluate the $n$-step cyclic dynamics far enough ahead. This however is expensive to compute.

We thus extend the vanilla CEM model to make the relationship between a character’s health and their performance more consistent with nature. More specifically, we assume that a character’s actions are harder to control the more they are injured by introducing noise into the character’s environment dynamics model. In the following, we define a successful action as one that results in the follow-up states $\hat{r}_{t+1}$ expressed in the original dynamics $q(\hat{r}_{t+1} \mid \hat{a}_t, \hat{r}_t)$. An unsuccessful action leaves the present state unchanged. We introduce noise by setting the probability for a character’s actions to be successful in a state $\hat{r}_t$ proportional to the fraction of their remaining and maximum health:

$$q(\hat{r}_{t+1} \mid \hat{a}_t, \hat{r}_t) = \left[ q(\hat{r}_{t+1}^1 \mid \hat{a}_t, \hat{r}_t), q(\hat{r}_{t+1}^2 \mid \hat{a}_t, \hat{r}_t), \ldots, q(\hat{r}_{t+1}^D \mid \hat{a}_t, \hat{r}_t) \right] \odot \left[ \begin{array}{c} \gamma \\ \gamma \\ \vdots \\ 1 - \gamma \end{array} \right], \gamma = \frac{h_t}{h_{\text{max}}}$$

Figure 6.6: Experiment 3. The player is threatened by an enemy. The NPC could save the player by shooting the enemy or by stepping between them.
Figure 6.7: Experiment 3. Player empowerment, given the NPC chose to shoot in a certain position, and player health > n, with n = 2. Health-performance consistency provides a clear indication for the NPC to shoot.

Here, ⊙ is the element-wise product, and $h_t, h_{\text{max}}$ stand for the character’s current and maximum health, a representative for some arbitrary fitness label. The state $\hat{r}^D_{t+1}$ resembles the agent’s default follow-up state, i.e. the state resulting from idling. The more a character’s health decreases, the more likely it becomes that their actions will lead to the default state. Since empowerment punishes overlapping action consequences, this modification yields a consistent, gradual decrease of a character’s empowerment with their health, assuming that it is applied to all available actions.

When including this modification, a CEM-driven NPC not only acts when the player faces death, but also protects the latter from being harmed. Fig. 6.7 illustrates this in the previous scenario, by comparing the player empowerment landscape for the vanilla and extended formalism when the player’s current health $h_t$ is larger than the lookahead $n$. Again, the landscape has been computed for a fixed player and different positions of the north-facing NPC. Our health-performance consistency allows the NPC to clearly differentiate between actions that contribute to the player’s empowerment, despite a short lookahead $n$. Hence, we assume this extension to be active by default in the following scenarios. A video showing how the CEM-driven NPC protects the player can be found in the digital appendix.

One of the most frequently criticised shortcomings of existing NPC AI is that the controlled NPC blocks the player’s movement, and hence their goal achievement. Emmerich, Ring and Masuch (2018) have asked players what they like and dislike in companion characters, and, as an open response, players particularly complained about companions getting in their way. In our fourth experiment, we probe if a CEM-driven NPC would block the player’s movement while maintaining spatial proximity.

We also investigate this in a game level representing a single room. Fig. 6.8a shows the player’s empowerment for different NPC positions. The values are low around the player, because the NPC would constrain their movement. The same applies to the NPC’s periphery in respect to the player. If we
only considered these two types of empowerment, they would add up and lead to repellent behaviour; but in combination with NPC-player transfer empowerment, the coupled empowerment around the player increases (Fig. 6.8b), and the NPC consequently maintains spatial proximity, but avoids blocking the player whenever possible. A CEM-driven NPC consequently prefers to position themselves either in the corners in front or behind the player, where they do not present an obstacle. We have included a video of the NPC following but not blocking the player in the digital appendix.

We conclude that CEM enables NPCs to maintain the player’s integrity by protection against enemies and by not blocking them.

6.5.1.7 Duty 3: Ensure Own Integrity

We have demonstrated earlier that the companion protects the player from threats. Various earlier studies (e.g. Guckelsberger & Polani, 2014) have demonstrated that EM makes agents death-averse, which drives the NPC to defend themselves against threats. In our fifth experiment, we look at a dilemma addressing both companion’s duties to protect the player and themselves: when NPC and player are threatened at the same time.

We probe this behaviour with a similar condition as in the first experiment, extended by an additional enemy that threatens the NPC. We first consider the case where the player has sufficient health to withstand the enemy for several moves, rather than getting killed immediately. Fig. 6.9 illustrates the behaviour of our NPC in this scenario as a series of movements. The first image in the series shows the coupled empowerment in the initial situation. Here, the dark area between the companion and the enemy on the left renders the latter as a threat, while the white area towards the other enemy represents the companion’s potential to save the player from harm. The following images
Figure 6.9: Experiment 5. NPC and player are threatened simultaneously. Successive moves from top left to bottom right: the NPC escapes their death, rescues the player, and finally defends themselves. Arrows indicate shooting. Top left: coupled empowerment for $n = 2$ and different NPC positions.

demonstrate that the companion first escapes from the enemy on the left, while accepting that the player is harmed. It then kills the player’s enemy before the latter can attack the player. The remaining enemy follows the NPC until the latter eventually kills the enemy to save its own existence.

We moreover consider a second case, in which the player’s health is set to one from the onset, and they are hence immediately threatened by their enemy. For $\alpha_p > 0.5$ and the other hyperparameters set to default, the NPC sacrifices themselves to rescue the player. The digital appendix comprises a video with both scenarios. Our fifth experiment has demonstrated that a CEM-driven NPC can protect the player and themselves in complex behavioural sequences emerging from their social intrinsic motivation.

Our experiments support the notion that a CEM-driven NPC can yield believable and supportive behaviour towards the player, operationalised by the three duties to maintain operational proximity, and ensure the player’s as well as their own integrity. The NPC has realised these duties without a policy model that is custom-tailored to a specific player, and we thus consider CEM player-general. In conclusion, we confirm our predictions PD.1 and PD.3 for this testbed. In the next section, we study the potential of CEM to drive adversarial characters. We combine the discussion of both studies in Sec. 6.6.
6.5.2 Study 2: Adversary Non-Player Characters

In our second study, we investigate CEM’s capacity to drive the behaviour of general and believable adversary NPCs. To this end, our NPC essentially chooses actions that increase their own, and decrease the player’s empowerment. In analogy to our first study, we question whether CEM can sustain characterhood, here in terms of different facets of believable antagonistic behaviour (PD.2). Differently though, we evaluate the generality of these characters with respect to their ability to respond flexibly with new but consistently antagonistic behaviours to changes in their embodiment (PD.4) and environment (PD.5). We first clarify what we mean by adversaries and what facets of believable, adversarial behaviour we investigate in the following experiments.

6.5.2.1 Believable Adversary Non-Player Characters

NPCs can realise many different roles in videogames, but the majority of characters found in games are adversaries. They come in different flavours, with their roles and behaviours varying according to the game genre, the design affordances, and the underlying algorithms. The terminology used in the literature is incoherent and overlapping. Warpefelt (2016, p. 87-90), for instance, differentiates adversaries into enemies and opponents. While enemies are reduced to the roles of attackers in combat, opponents are described as more complex, acting against the player’s goal-oriented manipulation of the game mechanics. Instead of just attacking the player, they could hinder them ‘from moving, chase them, or force them to alter their plans’ (ibid., p. 90 ff). Treanor et al. (2015) in contrast distinguish adversaries and villains. While an adversary is conceived as a character that can defeat the player without resorting to cheating, a villain’s primary goal is seen in creating interesting challenges for the player which they can eventually overcome. In the following, we understand adversaries to be as complex as opponents. They either challenge the player as villains, or actually defeat them.

We have already foreshadowed the primary contribution of adversary NPCs to player experience (PX): they produce challenge as a central constituent of gameplay (Adams, 2014). Adversaries can challenge players physically through speed and accuracy, e.g. in battle, and cognitively by addressing their problem-solving capacities, e.g. in the form of puzzles or sophisticated strategies. Challenge serves as an anchor for other PXs, e.g. enjoyment, competence, suspense and curiosity, anticipation and tension, and of course, success and failure (Denisova, Guckelsberger & Zendle, 2017). If challenging them optimally, adversary NPCs can thus contribute substantially to player satisfaction and continuous engagement with a game.

The positive effects of adversaries on PX can be obliterated if they do not behave in a believable manner as conceived in Sec. 6.2. To realise characterhood as a central determinant of believability, an adversary must challenge the player through behaviour that is antagonistic with respect to the player’s own and the goals of the game. An adversary should moreover show diversity in behaviour both within a given game, and with respect to the player’s expectations shaped by other games. In other words, it should behave in non-stereotypical, new and surprising ways. This is promoted if an adversary
leverages their environment, in combination with their own abilities, to exploit the player’s weaknesses. Moreover, they should be attentive to the player: Warpefelt (2016) notes that in existing games, adversaries often lack believability as they appear ‘oblivious until actively provoked’ (ibid., p. 89). Both points are covered by the third requirement for believable NPCs to be sensitive to their surroundings and body. Finally, adversaries can appear more believable if they follow their own agenda, even if it is only about self-preservation expressed in defensive, rather than kamikaze behaviour. The design of believable adversary behaviour is subject to the same challenges as in designing believable NPC behaviour more generally, as outlined in Sec. 6.2. In complex, open-ended games, existing NPC AI may break or yield blunt behaviour. This impedes on the characters’ believability and thus on PX.

6.5.2.2 Method

As in our previous study (Sec. 6.5.1), we assess and explore our predictions via observational vignettes to gain qualitative data on the NPC’s behaviour. The player avatar is controlled by the experimenters. For PD.2, we probe the effect of CEM on two factors of believable adversary behaviour: their (i) realisation of characterhood, and (ii) sensitivity to their surroundings and body.

We operationalise (i) based on two facets of believable adversary behaviour: predator-and-prey dynamics and attacks from a distance. We have chosen the first for its omnipresence in many games, and the second as a more sophisticated and unconventional ‘means to be mean’. We have dedicated two parts of our study to probing each of these facets (Sec. 6.5.2.5 and 6.5.2.7). We investigate (ii) in a third part (Sec. 6.5.2.6), by observing the behaviour resulting from changes to an initial environment and set of character abilities. In our default setup, the CEM-driven NPC bases its decision-making on the maximisation of its own, and on the minimisation of the player’s empowerment to the same extent. We later deviate from this equilibrium and show how unbalanced configurations yield radically different behaviours and adversary personas. We probe (i) and (ii) via nine experiments in total.

We investigate the generality of our NPC by two means. Firstly, we observe whether they exhibit consistently antagonistic behaviour throughout all experimental setups. Secondly, we check whether they respond with new antagonistic behaviours to changes in their environment (PD.5) and abilities (PD.4). A character that can flexibly respond to such changes is also likely to exhibit behavioural diversity, and we hence implicitly address another facet of believable NPC behaviour.

Due to the richness of our extended testbed, an exhaustive search through the space of environment features and character abilities is infeasible. We hence focus on those combinations that yield the biggest difference in emergent behaviour, and that can be found in many games and thus capture the use-cases of game designers and researchers. We also cannot exhaustively evaluate all combinations of the lookahead $n$ and the weights $a$ in the coupled empowerment action-value function, and thus only study those that illustrate CEM’s sensitivity to, and the benefits of, parameter fine-tuning best.
### Testbed

We have adopted our dungeon-crawler testbed from the previous study on CEM-driven companion NPCs (Sec. 6.5.1.3) for its inherent benefits outlined earlier, to support comparisons, and to provide a basis for a future joint quantitative evaluation. Despite being rather minimalistic, this testbed allowed us to investigate an NPC’s generality in responding flexibly to the unknown behaviour of the player and other characters. In this study though, we investigate generality with respect to an NPC’s ability to respond with new antagonistic behaviours to changes in their embodiment and environment.

To this end, we have extended the testbed substantially with both new environmental features and character abilities. Tbl. 6.4 provides an overview of the various features, their dynamics and the rationale behind their inclusion. As a minor change and visual guide, we have now included a goal tile which the player must navigate to in order to win the game. Moreover, we have changed the NPC and player avatar (excluded from the table) to a circle, and now explicitly visualise their perceptive field. A summary of the extended character abilities can be found in Tbl. 6.5.

### Model

We essentially use the same model as in our study on companion NPCs (Sec. 6.5.1.4), but with a negative hyperparameter \( \alpha_F \) such that maximising coupled empowerment leads to the NPC performing actions that minimise the player’s empowerment. Moreover, we limit our investigation to the interaction...
of the NPC and player, and omit other characters. We adopt the health-performance consistency extension (Eq. 6.49) of the vanilla CEM formalism introduced earlier to make the NPC sensitive to their own and the player’s health even for low lookaheads $n$. We however do not use the trust correction extension, but assume policy models to be uniform distributions.

We adopt the previous assumptions that the characters’ sensors are asymmetric, local and non-overlapping, and that their perceptive field spans a radius of three cells. While the characters’ abilities were previously fixed, we now modify them across different experiments. We assume a consistent default configuration, in which they can only idle and move, i.e. $A = \{\text{idle}, \text{north}, \text{east}, \text{south}, \text{west}\}$. Furthermore, they are initialised with full health, i.e. $h_{t=0} = h_{\text{max}} = 2$. This allows them to take damage without dying instantly, and to make use of health rechargers. A character takes one health point damage if hit by another’s melee attack, a range attack or through a turret. Vice versa, they gain one health point if healed by another character. If subjected to lava, they lose one health point per time step, and gain one while standing on a health recharger. The remote attack range is four tiles.

We compute coupled empowerment for a 3-step lookahead, and assume an initial weighting of empowerment types based on the hyperparameter values $\alpha_C = 0.5, \alpha_P = -0.5$ and $\alpha_{CP} = 0.1$. These values have been identified through experimentation. We only report on these settings in our experiments if they deviate from the default configuration.

<table>
<thead>
<tr>
<th>Ability</th>
<th>Description</th>
<th>Reason for inclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Idle</td>
<td>Causes no change to the current game state.</td>
<td>Can serve as a fallback if other actions are disadvantageous.</td>
</tr>
<tr>
<td>Move</td>
<td>Move character into adjacent cell if there is no obstacle. Otherwise only changes orientation.</td>
<td>Allows for exploration, hiding and change of position in reaction to other characters.</td>
</tr>
<tr>
<td>Push</td>
<td>In addition to moving, shift adjacent characters in the movement direction if there is no obstruction.</td>
<td>Allows for complex interactions, e.g. pushing others into lava, rechargers, or a turret’s target range.</td>
</tr>
<tr>
<td>Fly</td>
<td>Allows to move over lava fields without taking damage, but character can still benefit from rechargers.</td>
<td>A way to access previously inaccessible parts of a level, and escape other characters.</td>
</tr>
<tr>
<td>Melee attack</td>
<td>Causes damage to faced, adjacent. The amount of health damage is predefined and fixed.</td>
<td>Common predator-and-prey mechanic. Allows for escape or kill with range attack before others close.</td>
</tr>
<tr>
<td>Range attack</td>
<td>Reduces health of first character in current direction and attack range. Damage and range are predefined.</td>
<td>Allows experimenters to imbalance attack options based on spatial proximity. Rewards seeking cover.</td>
</tr>
<tr>
<td>Heal</td>
<td>Increases health of adjacent, faced character by fixed amount up to maximum health of that character.</td>
<td>To check how an unconventional action can benefit adversary-player interaction.</td>
</tr>
</tbody>
</table>

Table 6.5: Extended character abilities in our study of adversary NPCs.
6.5.2.5 Facet 1: Predator-and-Prey

As the first part of our study, we investigate whether CEM can give rise to the classic predator-and-prey behaviour. This is quintessential to many games, and represents our first facet to operationalise believable adversary characterhood. Fig. 6.10a shows the initial state of the environment, consisting of an arena surrounded by walls, and divided by a wall with small spaces on the sides to pass through. The NPC (‘A’, orange) is at the top and faces south, while the player (‘P’, purple) is situated at the bottom and faces north. Their perceptive field is highlighted in orange and purple, respectively.

Fig. 6.10b shows the NPC’s empowerment for a 3-step lookahead if they were moved to that specific position, but the player remained fixed. The NPC itself is hence omitted from the empowerment landscape. Brighter hues represent higher empowerment. In the default configuration, the characters can only move or idle, and empowerment is consequently very sensitive to degrees of freedom in movement: it is lower where the NPC would be blocked, e.g. close to walls, corners and the player. The player’s 3-step empowerment is very similar as both characters by default possess the same abilities, and we hence do not illustrate it. Fig. 6.10c shows the NPC-player transfer empowerment for different NPC positions. Recall that this empowerment type corresponds to the influence the NPC has on the player’s sensor. Hence, for \( n = 1 \), it is only non-zero within and directly adjacent to the player’s perceptive field. For larger lookaheads in contrast, it fades out to states from which the NPC could influence the player’s perception with an \( n \)-step action sequences (Fig. 6.10d). This demonstrates that transfer empowerment does not measure perceptibility, but operational, or in this case, spatial proximity.

These illustrations invite predictions on how the different empowerment types compete in the CEM policy: If the NPC based their action-selection only on transfer empowerment, they would move closer to the player; maximising NPC empowerment as their own agenda however would require them to stay in the middle of the upper part and avoid the choke points on the sides. As this trade-off is mediated by the \( \alpha \) hyperparameters, they can be used to design different behaviours. For our first simulation experiment, we equip our NPC with the ability to perform range attacks but stick to the default parameter setup. We find that the NPC remains in the upper area, but once the player moves into this territory, they are killed with two precise shots. We consider this opportunistic predator-and-prey behaviour.

For our second experiment, we decrease the weight of the NPC’s own empowerment but increase the negative weight of the player’s \( (\alpha_C = 0.1, \alpha_P = -1.0) \). The NPC hence more strongly trades off losses in their own empowerment for the decimation of player empowerment. As a result of giving their own agenda less consideration, the NPC becomes a daredevil adversary, chasing and shooting the player throughout the whole level as illustrated in Fig. 6.11.

As a third experiment, we probe our NPC’s generality by also allowing the player to perform range attacks. For the default parameter configuration, the NPC still behaves opportunistically, but now dodges the player and keeps distance to avoid being attacked. The digital appendix contains three videos of the described NPC behaviour in the previous experiments.

Experiment 1: Opportunistic

Experiment 2: Daredevil

Experiment 3: Dodging Player Attacks
These three experiments supports the notion that CEM can yield adversary behaviour in the form of the classic predator-and-prey dynamics present in many games. They show that the \( \alpha \) parameters in the CEM action-value function should not be considered a burden, but a feature to create different NPC personas, thus increasing the believability and diversity of our characters.

6.5.2.6 Facet 2: Sensitivity to Surroundings and Body

In a complex game, the wealth of possible interactions between a character’s abilities and features of the environment becomes hard to anticipate even for the game’s designers (cf. Sec. 6.2). As a consequence, most traditional NPC AI does not fully exploit these interactions. For more advanced techniques such as Monte-Carlo Tree Search or RL, this anticipation problem creeps into the definition of the optimisation objective, resulting in blunt adversary
Empowerment is defined on an agent’s possible interactions with their world (cf. Sec. 3.2), and should thus be sensitive to any interaction between any type of ‘functional content’ (Smith, 2014b). As a second facet of believable adversary behaviour, we probe whether a CEM-driven NPC is sensitive to their surroundings and body, and can fully leverage the possible interactions that a game affords for adversarial behaviour. Since even small changes to the environment and a character’s abilities can turn the emerging gameplay around, we start with a simple scenario and modify it gradually. In this way, we also probe CEM’s generality with respect to such changes.

Fig. 6.12a shows the initial state of the environment for our fourth experiment: NPC and player face each other in an arena surrounded by lava. Stepping on lava reduces a character’s health by one unit per time step. To examine longer interaction sequences, we extend our characters’ health to four units ($h_t = h_{\text{max}} = 4$). Mediated by health-performance-consistency (Eq. 6.49), a decrease in health results in lower empowerment even for small lookaheads. The NPC’s 3-step empowerment (Fig. 6.12b) is thus lower in the lava, and decreases further away from the platform, where only few action sequences lead back alive. Under the default configuration, the NPC gets close to the player and blocks them, to reduce their mobility and hence empowerment.

However, if we change the NPC’s abilities, the dynamics change considerably. For our fifth experiment, we equip them with the ability to push. As illustrated in Fig. 6.13, the NPC then kills the player by pushing them into the lava. Importantly, they block the player from returning to the platform, no matter which path the latter chooses. The policy thus captures how the NPC’s new ability, in interaction with the environment, can be exploited to
6.5 STUDIES: CEM FOR GENERAL, BELIEVABLE & SOCIAL NPCs

Figure 6.12: Lava arena scenario. Initial game state with perceptive field and 3-step NPC empowerment for different NPC positions and a fixed player position. The illustrations show the impact of different modifications to the environment and character abilities on empowerment.

decrease the player’s empowerment. We have included a video of this more challenging and arguably novel behaviour in the digital appendix.

In our sixth experiment, we consider the consequences of giving the NPC an action which is typically not associated with adversarial behaviour: to heal other characters. Surprisingly, this change takes antagonism to another level: the NPC still pushes the player into the lava, but once they are close to ceasing, the NPC uses healing to keep them barely alive. Our CEM-driven NPC thus acts in a super-villain style, and in contrast to e.g. an RL approach
that only rewards the player’s destruction: they keep the player’s health just high enough to exercise control over them – thus optimising their own empowerment and NPC-player transfer empowerment. We can modulate this behaviour via the hyperparameters: if we reduce $\alpha_C$, the NPC lets the player die. We include a video of the healing behaviour in the digital appendix.

An NPC cannot sustainably maximise coupled empowerment if they exclusively exercise control over the player; in scenarios such as the lava arena, they must also engage in acts of self-preservation, thus realising an agenda of their own. Dodging attacks by the player as in the third experiment is such an act. However, previously both characters had identical abilities, which is uncommon for most games. We hence set up a seventh experiment to examine whether CEM can exploit inequalities between characters to further both self-preservation and adverseness. To this end, we allow our NPC to range-attack and fly, while the player is limited to melee attacks on the ground. In our testbed, a character that can fly is not affected by the hazardous effect of lava, and the NPC’s empowerment is thus not affected by the lava anymore, but only by the surrounding walls and the player. This can be seen by comparing Fig. 6.12b and 6.12c. With their new ability, the NPC dodges the player’s melee attacks by escaping over the lava. Once the player veers away from the NPC, they return and attack from a distance. We have included a video of this behaviour in the digital appendix.
In our eighth experiment, we probe another aspect of self-preservation: not escaping harm, but recovering from it. If we allow our characters to push and perform melee attacks, the NPC engages in close combat, using both direct attacks and pushing the player into the lava. Meanwhile, if we place a health charger unit in the middle of the platform, the characters start competing for the scarce resource: once the NPC’s health gets close to zero, they capture the recharge tile to recover, pushing the player off the tile if necessary. If we set the NPC’s health lower than their maximum, i.e. $h_{1=0} = 2$, $h_{max} = 4$, the health charger appears like a beacon in the NPC empowerment landscape (Fig. 6.12d). The digital appendix contains a video of the emerging behaviour.

The preceding experiments support both, that CEM-driven agents are sensitive to their surroundings and body as a facet of believable behaviour, and that they are general with respect to changes to these elements.
6.5.2.7 **Facet 3: Distant Attacks**

The most challenging adversaries arguably strike from a distance, where they remain unaffected by the player’s actions, and potentially undetected. An **NPC** that is inferior in direct combat could cast spells, order air strikes or control traps remotely. In this last study part, we investigate if **CEM** can yield such attacks from a distance as a facet of believable adversary behaviour.

We have designed our last experiment to provoke a transition between a remote and direct attack, and to examine the interplay of the three rewards in the **CEM** action-value function. In the initial state (Fig. 6.14a), the player starts on the lower right in a corridor, while the **NPC** is situated on the upper left in an open area, separated by a wall with two passages. The player faces three **turrets**, two on the sides and one ahead. The corresponding **triggers** are positioned in front of the **NPC**. Both characters can perform a **range attack**.

The **NPC**’s own empowerment in this state does not convey any information about the best trigger to affect the player, as it only quantifies their influence on their future own sensor state. Player and **NPC**-player transfer empowerment in contrast both work as a proxy to the player’s condition: transfer empowerment measures the impact of the **NPC**’s turret-triggering on the player’s health, which is captured in the latter’s sensor; the player’s health in turn affects their empowerment, which can be exploited by the **NPC**. Figs. 6.14b – 6.14d show how transfer empowerment peaks on and around the triggers for player positions in the shooting range of different turrets.

Following the **CEM** policy, the **NPC** triggers the correct turrets to hit the player on their way towards the goal tile (Fig. 6.15, step one to four). When the player moves between turrets, the **NPC** positions themselves where they can strike quickest, i.e. between the triggers. Once the player gets closer to the goal and thus to the open passage towards the **NPC**, the latter trades...
off their own empowerment, and their NPC-player transfer empowerment: the spatial proximity of the player results in a transfer empowerment gradient which the NPC could follow to eventually attack the player directly. By doing so however, the NPC risks their own empowerment to be decreased by a counter-attack. In the present configuration, the NPC eventually moves away from the triggers and attacks the player directly (Fig. 6.15, step five).

Meanwhile, decreasing the NPC’s health \( h_0=1, h_{\text{max}}=2 \) makes it remain at its current position and shoot the player from a distance. We include two videos covering both modifications in the digital appendix. This experiment supports the claim that CEM can also produce more sophisticated forms of antagonistic behaviour.

Our study shows that minimising the player’s empowerment in a CEM policy allows an NPC to challenge the player with believable, antagonistic behaviour and thus realise the role of an adversary. We have operationalised believable antagonism based on the classic predator-and-prey behaviour found in many games, and sophisticated attacks from a distance. As a third facet of believable behaviour, we have probed whether the NPC remains sensitive to their surroundings and body. Without modifications to the underlying model, they have consistently responded to changes in their environment and embodiment with new and often surprising antagonistic behaviour. For this testbed, we hence confirm our predictions that CEM can realise environment and embodiment-general (PD.4, PD.5), believable and antagonistic (PD.2) behaviour. We next conclude this chapter with a discussion of our two studies.

We have introduced social models of IM to increase the generality of artificial agents in co-creative interaction, while constraining their behaviour to either support or antagonism. We have instantiated this proposal in CEM for application in videogame AI to drive the behaviour of NPCs as co-creative agents. Across two qualitative studies based on observational vignettes, we have explored five predictions of the capacity of CEM to drive the behaviour of believable NPCs that either support (PD.1) or challenge (PD.2) the player as companions and adversaries, while remaining general with respect to the specific player (PD.3), and with reference to changes in their embodiment (PD.4) and environment (PD.5). We find that CEM, subject to minor modifications to the original formalism, fulfils all predictions within the limits of our dedicated game testbed. Here, we briefly reflect on the core findings of our studies. We then discuss limitations of these studies, and restrictions of the underlying CEM model in its present form. We finally draw on our findings to re-evaluate how CEM could benefit CC and other fields.

Reflecting on our study, we want to stress three central findings. Firstly, we managed to drive NPCs to either support or challenge the player with essentially the same underlying principle – we only had to flip a single hyperparameter, but left the rest unchanged. In contrast to established NPC AI, we did not engineer a small set of behaviours that repeat over time, or designed an extrinsic reward landscape that only works on a specific game. Rather
than designing a character to populate a specific world, we created a world that gave rise to their behaviour, emerging from the possible interplay of the NPC’s abilities, the abilities of other characters, and the elements of the environment. There was no need for us to update the model with semantic annotations of new game elements; it only relied on the interaction between these and other game elements to be encoded in the forward model.

Secondly, our NPCs have exhibited player-generality (Togelius & Yannakakis, 2016) in consistently supporting or challenging the different experimenters without modelling their specific policy. Moreover, they have shown limited game-generality (ibid.) by maintaining these social dynamics in response to changes in their embodiment and environment.

Thirdly, variations of the hyperparameters yielded different, surprising NPC personas such as an ‘opportunist’, ‘daredevil’ and ‘super-villain’. Contrary to the typical design process, we have discovered what types of characters a given game world affords, rather than designing them for that world.

Our study provides the foundation for the further investigation of CEM. Crucially though, it also has several limitations which must be overcome for a full proof-of-concept and practical application of CEM-driven NPCs.

Observational vignettes as exploratory and qualitative study method allowed us to describe the nature of CEM-driven behaviours in rich detail, classify them into the somewhat ambiguous categories of support and antagonism, and identify the formation of different personas. Due to its qualitative nature though, the vignettes could not reveal the strength of CEM-induced support and antagonism as perceived by a player or designer. This information however is vital for designers to afford a good player experience through sufficient support and an optimal level of challenge. Moreover, we have noted the novelty and surprisingness of the exhibited behaviour, but we could not quantify it. This however would be desirable for the formal assessment and comparison of an NPC’s believability through the determinants of characterhood and behavioural diversity (Tbl. 6.1 and Sec. 6.2).

Our minimalistic but highly versatile, dedicated game framework enabled us to probe specific facets of believable NPC behaviour via custom-made conditions, and granted us maximum insight into the characters’ motivation through the calculation of reward landscapes, amongst other means. Crucially though, it only affords limited insights on the game-generality of CEM. While we have applied the same model to different scenarios, they are all situated in a small space of possible game mechanics, and only represent a single genre. Moreover, our testing method could be misinterpreted as ‘cherry-engineering’ only those scenarios that give rise to the desired believable behaviours.

We chose to steer the player avatar ourselves to playfully explore the possible interactions with our CEM-driven NPCs in different worlds and for different hyperparameter configurations. However, this approach limits our insights on CEM’s player-generality: not only were the NPCs only exposed to the playing styles of few experimenters, but there is also the risk of them introducing an unconscious bias into their gameplay to provoke certain desirable behaviours, while avoiding others. We propose means to overcome these limitations of our present study as part of future work in Ch. 8.
We moreover consider limitations to CEM. We distinguish fundamental restrictions to its game-generality, as well as temporary limitations to its practical application in commercial games resulting from its present formulation.

CEM can realise believable NPC support or antagonism only in games that require players to either increase or decrease their options and influence in order to progress towards the game’s goals. In other words, CEM’s game-generality is conditional on a positive or negative implicit goal alignment (Sec. 5.1.4) with empowerment, providing a gradient for CEM to operate on.

If a game expresses such an empowerment gradient towards goal achievement, it typically is not a constant slope but a bumpy road that provides challenge and enjoyment: we may only be able to progress into a different part of a *Doom* (id Software, 1993) level by facing a hoard of enemies in a connecting room. Similarly, we may have to defeat a minor enemy in *StarCraft* (Blizzard Entertainment, 1998) to prepare for the final match with a much stronger opponent. In each case, we temporarily endure situations in which we have fewer perceivable options, captured by empowerment, in order to multiply them later. Two distinct elements of the CEM formalism put the NPC at risk of getting stuck into local coupled empowerment maxima: the greedy action-selection function, and the specific choice of model hyperparameters, including the lookahead, which shape the reward itself. In our experiments (Sec. 6.5), we fixed these hyperparameters to a default configuration which produced sensible behaviour throughout all conditions, but this is likely not universally applicable. In Ch. 8, we discuss future work on choosing an appropriate action-selection function and on solving the hyperparameter setup as means to increase CEM’s game-generality.

We next discuss present limitations of CEM with respect to its practical application. A game is more than the sum of its parts (Liapis, Yannakakis & Togelius, 2014), and CEM is sensitive to this complexity. The behaviour of a CEM-driven NPC emerges from the interaction of the specific parametrisation of the action-value function, embodiment, environment and interaction partners. If only one of these components changes slightly, the emerging behaviour – within the boundaries of support or antagonism – can shift substantially, and is hence hard to anticipate. This has been the motivation to develop the principle in the first place (Sec. 6.2): by relying on IR, CEM promises to yield sensible behaviour in response to such changes. The flip side of this increased robustness and potential for diversity and generality is a lack of predictability, as the emerging behaviour also exceeds what a designer can anticipate. In principle though, we could predict the behaviour of a CEM-driven NPC, if it was instantiated with a deterministic policy and all elements in the game world as well as their interaction were known. Predictability is thus only limited by human cognitive bounds, rather than by an element of chance. This yet threatens CEM’s practical application: Yannakakis and Togelius (2018, p. 14) point out the games industry’s resistance towards embracing NPC AI whose behaviour is not entirely predictable at design time. We believe that CEM can only be fully leveraged and benefited from if this resistance to uncertainty is overcome. Nevertheless, we still propose means to accommodate this industry requirement as part of future work.
Another major industry requirement towards NPC AI is efficiency. To discriminate the effect of small changes to the CEM model on NPC behaviour in our studies, we have calculated the simplified coupled empowerment exhaustively as in Alg. 1. In the following complexity estimate, we ignore the calculation of NPC-player transfer empowerment. We assume full observability, and that NPC and player have the same number of actions \( |A^p|=|A^c|=|A| \). The time complexity is a function of the number of actions \( |A| \), the lookahead \( n \), and a branching factor \( k \), with \( 1 \leq k \leq |R| \). The latter represents the number of environment states that an action leads to on average. We distinguish two parts of the overall coupled empowerment reward calculation.

Firstly, the computation of the NPC and player empowerment requires calculating the \( n \)-step cyclic dynamics for the \( n \)-step predictive factor. If one interaction cycle only comprised the interaction of the NPC and player, and they both used a uniform model of their peer’s policy, the NPC must perform \((|A^c|^2)^n\) calls on the respective environment dynamics models for each predictive factor. This must be computed for each state in which the player and NPC can act next, i.e. for \((|A|)^k\) states at \( t_p \) and \((|A|^2)^k\) states at \( t_c+\tau \). The calculation of all predictive factors hence has a time complexity of \((|A|)^k(|A|^2)^k + (|A|^2)^k(|A|)^{2n} = (|A|^2)^{2n+1} + (|A|)^{2n+2} \). Asymptotically, we have a time complexity of \( \text{O}_{\text{Pred}} = O((|A|)^n) \).

Secondly, we must calculate empowerment for each predictive factor. Calculating empowerment exhaustively for a single factor based on the Blahut-Arimoto algorithm (Arimoto, 1972; Blahut, 1972) has a time complexity of \( |A|^n|R_*| \). Here, \( |R_*| := (|A|^2)^n \) represents the average cardinality of the predictive factor. Repeated for every factor, the overall time complexity is \((|A|^2)^{n+1}|A|^n(|A|^2)^n + (|A|^2)^{2n+2}|A|^n(|A|)^2n = |A|^n(|A|^2)^{4n+1} + |A|^n(|A|^2)^{4n+2} \). Asymptotically, we have \( \text{O}_{\text{E}} = \text{O}(|A|^n(|A|)^n) \).

The overall time complexity of the coupled empowerment reward calculation results from summing the complexities of the \( n \)-step predictive factor and the empowerment calculation, and is exponential in the NPC’s lookahead. It also depends on the number of available actions and their branching factor, but an efficient implementation of the model makes it independent of the overall state space size \( |S| \). A CEM-driven character with fixed lookahead could thus be employed in a more sophisticated game world without increasing computational complexity, as long as it does not afford characters more actions, or branches these actions more widely.

The branching factor \( k \) is 1 for deterministic environment dynamics, and typically small in the stochastic case. Moreover, the NPC’s and player’s action outcomes often overlap, thus reducing \( |R_*| \). Even so, CEM can at present only be computed for relatively small lookaheads. This limits the NPC’s behavioural complexity, as the coupled empowerment can only reflect the impact of near-future events on the individual characters’ potential and perceivable influence. They would only recognise a time bomb as a threat, once its counter falls below their lookahead. CEM’s practical application is...
also limited because commercial games typically only provide a small budget for the computation of NPC AI. We discuss avenues to optimise both parts of the computation and scale CEM up for commercial application in Ch. 8.

Our study so far only demonstrates the game and player generality (Togelius & Yannakakis, 2016) of CEM with respect to a small set of game variants, and a non-representative sample of players. We elaborate means to provide stronger evidence on the principle’s capacity for generality in Ch. 8. Here, we want to explicitly note that our health-performance consistency extension in Eq. 6.49 does not fundamentally limit CEM’s generality; it is merely a short-cut to make the NPC sensitive to events that lie beyond their lookahead. This is not a necessity, and could be compensated for with a scalable approach to calculating the IR. Trust correction as our second extension in Eq. 6.48 does not impede generality either, as it has been specifically designed to rely on empowerment rather than an extrinsic reward.

We finally consider our study findings through the lens of CC to reflect on human-computer co-creativity in the interaction of a player and NPC, hence bridging between our CC and game AI motivation in Sec. 6.1 and 6.2, respectively. Our CEM-driven NPCs have exhibited novel behaviours that have not been explicitly programmed, but emerged from the interaction of the coupled empowerment intrinsic value function, the NPC’s embodiment, environment and their interaction partners. Even in our roles as researchers, and informed by our theoretical understanding of CEM, some emergent behaviours took us by surprise: we for instance neither expected the NPC in our lava scenario (Sec. 6.5.2.6) to block the player’s way back to the platform by following them along, nor did we expect them to heal the player and hence increase the perception of their antagonism. To call this behaviour creative (Runco and Jaeger, 2012; and Sec. 4.1.1), it is typically required to also have value. By distinguishing different interpretations of value, we identify three types of co-creativity in our study, one of which we have not anticipated ahead.

For an interaction to qualify as co-creative, the human partner and NPC must contribute to shared goals (Sec. 6.1), and we consequently link a person’s valuation of NPC behaviour to the fulfilment of such a goal. The first (i) identified type of co-creativity is characterised by both NPC and player contributing to the game’s goals. This is only realised by our companion NPCs (Sec. 6.5.1), whose supportive behaviour is valuable for the player in that it contributes to their achievement of the game. As example, consider our fifth experiment in Sec. 6.5.1.7, where the NPC helps the player to defeat their enemies and even sacrifices itself, hence ensuring the player’s survival as the dungeon-crawler’s implicit goal. The second (ii) observed type of co-creativity is realised in both our studies. Here, the shared goal is the player’s experience of interacting with a specific type of NPC. This requires the player to probe the NPC’s characterhood, and the NPC to enact it without violating the player’s expectations. As example, consider our fifth experiment in Sec. 6.5.2.6, where the player must pass by the NPC to reach the goal, just to find themselves being pushed into the surrounding lava. When trying to move back on the platform, the player is blocked by the NPC, and thus experiences another convincing facet of their adversary characterhood. We have identified a third (iii) type of co-creativity through our discovery of different NPC personas (Sec. 6.5.2.5 and 6.5.2.6). Here, the CEM-driven NPC contributes...
to the designer’s goal to create enjoyable and believable characters. In our role as designers, we have changed properties of the game environment, characters and NPC AI; the NPC in turn has performed a series of behaviours that informed our design of the next condition, which yielded another persona. At development time, such acts of mixed-initiative co-creation (Yannakakis, Liapis & Alexopoulos, 2014) could inspire new character designs, e.g. via changes to the game that enable more diverse or unorthodox behaviours. We suspect that, by surprising designers with unexpected behaviours, CEM-driven NPCs can provoke transformational creativity (cf. Grace & Maher, 2015).

We have applied CEM to drive and examine human-computer co-creativity in videogames. CEM is independent of a specific domain and could hence be employed in other co-creativity scenarios where an increase or decrease in options and influence is perceived as support or antagonism by the human partner. In Ch. 8, we propose future work to advance central goals of CC based on our work on NPC AI. Moreover, we discuss next steps to bringing social models of IM and CEM to other creative domains.

Based on our application and evaluation of CEM in NPC AI, we affirm this chapter’s research question, ‘Can we use a model of intrinsic motivation to engineer general and social co-creative agents?’ (RQ.8), within the limitations of our studies. We thus contribute to the overarching questions of this thesis by demonstrating that CEM as a new model of IM can advance videogame AI (RQ.2), and address core concerns of CC research (RQ.1).

CEM combines different types of IR to directly drive the process of a CC system. In the next chapter, we complement this motivational approach by contributing a novel use of IR to evaluate the product of such a system, which indirectly determines its overall behaviour.
In this chapter, we introduce a novel approach to predicting the human subjective experience of interactive artefacts by means of computational intrinsic reward (IR). We apply our approach to predict players’ experiences of videogames as arguably the most popular type of interactive artefact. Our approach is designed to mitigate challenges in evaluating procedurally generated content as a core area of videogame AI, and a focal point of computational game creativity (Liapis, Yannakakis & Togelius, 2014; Ventura, 2016a). We conduct a qualitative study to explore the research question:

RQ.9 Can we use IR to predict people’s experience of interactive artefacts in a general and autonomous way?

This chapter contributes to answering the overarching research question RQ.2 by showing directly how IR can advance videogame AI. Moreover, our approach has the potential to advance the autonomous evaluation of artefacts in CC more generally, and hence indirectly relates to RQ.1.

We dedicate the first two sections to motivating our contribution from the perspective of CC and videogame AI. In Sec. 7.1, we highlight the system-side evaluation of artefacts as a central requirement to advancing core CC goals. Moreover, we identify present challenges in estimating people’s subjective experiences, in particular of interactive artefacts. In Sec. 7.2, we reassess these challenges in the evaluation of videogames. We focus on the task of predicting players’ experiences of procedurally generated content, and identify shortcomings of existing work. We address these with a novel approach to predicting player experience (PX) through IR assessed on simulated AI gameplay, introduced informally in Sec. 7.3. We instantiate this generic proposal in Sec. 7.4 in the form of an empowerment-based player experience prediction (EBPXP) model. We formalise the model, and provide pseudocode for the experience prediction. As a first step towards validating our proposal, we explore which experiences EBPXP can potentially predict through a qualitative study on a custom-made game in Sec. 7.5. In Sec. 7.6, we contextualise the identified experiences in games user research, highlight the limitations of our study and model, and discuss how our findings can inform the application of our approach to other CC domains. All parts except the first have been published by Guckelsberger et al. (2017), in a more condensed manner. We particularly describe the algorithm, study setup and results in more detail.

The main contribution of this chapter is a novel approach to predicting player experience of game content, and its instantiation in EBPXP. We moreover contribute to CC by not only identifying open challenges in estimating people’s subjective experience of artefacts, but also by pointing out their impact on the creativity, creative potential and autonomy of CC systems, as well as on the system users. We contribute to videogame AI by uncovering how existing approaches to predicting PX constrain the generality of game AI and limit the potential of procedural content generation. For a detailed
account of how the work presented in this chapter relates to existing research on models of IM in CC and game AI, see Sec. 4.2.3 and 5.2.2, respectively.

### 7.1 Evaluating the Experience of Interactive Artefacts

According to the *standard definition of creativity* (Runco and Jaeger, 2012; Sec. 4.1.1), a creative product should be both novel and valuable. An individual could create such a product by accident; however, it is commonly agreed that human creatives can not only *generate* artefacts, but also *evaluate* them with respect to these and potentially other properties. Identifying such *generation* and *evaluation* in psychological theories of the creative process is not straightforward. Craveirinha, Barreto and Roque (2016) e.g. uncover them in Csikszentmihalyi’s (1997) sequential, five-phase model of creativity. They identify *preparation, incubation and insight* as generative acts, followed by the *evaluation* of the generated product, and its potential further *elaboration*.

Researchers agree that CC systems must likewise accommodate both *generation* and *evaluation*. This position has not only been shaped by an interest in modelling human creativity or in designing artificial systems that appear more creative, but also by the need to filter a system’s produced artefacts to not overburden the human user. The earliest mention of these two components is arguably in a 1993 study on computational letter design, for which McGraw and Hofstadter relate to the iterative process of guesswork (generation) and evaluation as ‘the central feedback loop of the creative process’ (ibid., p. 16). In the *creativity tripod*, Colton (2008) captures three necessary conditions for an artificial system to be considered creative: *skilfulness*, *imagination* and *appreciation*. While the first two concern the generation of artefacts, the latter incorporates the evaluation of value and novelty, respectively. For her *standardised procedure for evaluating creative systems*, Jordanous (2012) has empirically identified 14 key components of creativity, amongst which are *thinking* and *evaluation* as well as the *generation of results*. The *creative systems framework* (Wiggins, 2006a, 2006b) as a formalisation of Boden’s (2003) model of creativity (cf. Sec. 4.2.2) allows for the inclusion of evaluation rules to influence the traversal and hence generation of concepts. Pérez y Pérez (2007) distinguishes the generation and evaluation of ideas in the *engagement-reflection* model of creativity, and Ventura (2017) inscribes both generation and evaluation in a blueprint to building CC systems. The ability to evaluate their products and process is considered key to advancing CC systems beyond ‘mere generation’ (Ventura, 2016b), and a prerequisite for creative autonomy (Jennings, 2010). As such, evaluation represents a stepping stone towards achieving a central CC goal: to engineer artificial systems that can be considered creative *in their own right* (Colton, 2008).

In this chapter, we concentrate on *evaluation* rather than generation. More specifically, we focus on the *formative* (Karimi et al., 2018) assessment of artefact value during the creative process, in contrast to the *summative* (ibid.) assessment afterwards. From the *engineering perspective* (cf. Veale, Cardoso and Pérez y Pérez, 2019; Pérez y Pérez, 2018) and Sec. 4.2.1, CC systems...
are primarily designed to produce artefacts that can be appreciated by and hence benefit people. To be perceived as creative agents in their own right or to be accepted as equal partners in co-creativity, these systems must be able to predict how their artefacts would be evaluated by their human users or co-creators, without involving them explicitly. This can be complemented by a value assessment from the system’s own perspective.

The human evaluation of an artefact can be shaped by objective criteria, e.g. its utility, but also by subjective experience, e.g. in the form of a person’s aesthetic judgement. Modelling the latter represents an ongoing challenge in CC. Early on, McCormack (2005) highlights the open problem to measure ‘human aesthetic properties of phenotypes’ in a ‘machine representable and practically computable’ (ibid., p. 432) way to introduce efficient and independent evaluation into evolutionary music and art systems. In the meantime, this challenge has been mitigated in many CC domains and for different types of systems. Mexica (Pérez y Pérez, 2015b) for instance, an allegedly autonomous computational storytelling system, assesses the interest of a generated story based on modelling the development of tension resulting from the interaction of characters with different emotional links. Being strongly informed by human storytelling, the model requires an interesting story to have an introduction, a climax and a resolution. In the domain of visual art, the Drawing Apprentice (Davis et al., 2014) as an alternating co-creative agent identifies the recent behaviour of its human partner with one of three perceptual layers, and contributes to the drawing via the same layer to produce ‘artistically valuable’, co-creative (ibid.) outcomes. Each layer has a different perceptual granularity and is psychologically grounded, e.g. in Gestalt theory (Arnheim, 1965). Ekárt, Sharma and Chalakov (2011) model the aesthetic preference of their user in a partially interactive evolutionary art system. They initially fit a set of formal aesthetic measures based on a user’s explicit selection of preferred evolved images. Once fitted, the measures are used to automatically evolve images that appeal to the individual user. The previous examples rely on a specific theoretical model of aesthetics; in creative adversarial networks, Elgammal et al. (2017) combine this theory-driven with a data-driven approach to autonomously generating visual art. Similar to Saunders and Gero’s (2001) work, their approach is inspired by psychological theories which propose arousal as a determinant of the aesthetic experience, and which identify a drive against habituation in human art practice (Martindale, 1990). Their model is an extension of Goodfellow et al.’s (2014) generative adversarial networks, and yields images that are expected to arouse their viewers through their resemblance with human art, but cannot be classified in existing style categories. As their discriminator is trained on a large corpus of canonical paintings, it approximates the human aesthetic judgement.

Despite these advances, many CC systems still miss an evaluation of artefact value, and in particular do not estimate human subjective value.

1 Colton et al. (2020) contrast this dominant design focus with the use of CC to enable artificial systems to express something about their unique ‘machine condition’.
2 Note that a model of the human co-creator’s value assessment does not only benefit supportive co-creativity, but can also inform how to challenge the human partner best (cf. Sec. 6.1).
3 We argue elsewhere (Guckelsberger, Salge & Colton, 2017) though that no existing CC system so far can be said to genuinely model value from an agent’s own perspective.
Kantosalo and Toivonen (2016) observe in regards to co-creativity: ‘where generation is often held as the forte of the computational agent, evaluation then again is very much held as the domain of the human author’ (ibid., p. 83). In the context of music and language generation, Pearce and Wiggins (2012) point out that evaluation functions to identify ‘high-quality artistic structures’ are ‘an open research topic, partly because current models tend to be incomplete representations of the phenomena they capture but also because quality criteria are subjective and context-dependent’ (ibid., p. 643). Considering this and McCormack’s (2005) appraisal jointly with present work, the evaluation of subjective value presents an ongoing challenge.

In this chapter, we focus on estimating human subjective artefact value, and mitigate three shortcomings of existing models. Firstly, most existing models are (i) subject-unspecific: they quantify ‘subjective experience’ generically across subjects, but remain agnostic with respect to individual differences that may contribute to the experience of a specific person. In particular, most models are not sensitive to how a person’s experience is shaped by their, embodiment and situatedness, yielding a unique perspective on a specific artefact. Secondly, many existing approaches are (ii) context-specific. Theory-driven approaches are usually very narrow; even within a single creative domain such as visual art or poetry, a specific theory of e.g. aesthetics may not be universally applicable. Similarly, people’s reports on their subjective experience, as used in data-driven approaches, often do not scale beyond the context in which they were originally assessed. Thirdly, existing models (iii) are strongly dependent on people. This may sound paradoxical with regards to models of subjective experience, and hence needs more elaboration. For a CC system to unleash their full creative potential in acts of transformational creativity (Boden, 1990/2003; Wiggins, 2006a, 2006b), they must be able to self-transform by perceiving and changing their generator component (Linkola et al., 2017). However, due to their context-specificity, the resulting new content may render the present estimate of human subjective value inaccurate, as this estimate was based on the previously accessible area of artefact space. Each transformation of the generator, either induced by the system or a person, may thus require theory-based models to be re-tuned by their designer, and data-driven models to be re-trained with new human experience data. The consequences are threefold. If the evaluation function is not adjusted, it may present the user with artefacts that do not correspond to their perception of value in the transformed space. This might de-escalate to a degree where the system’s evaluation appears random, rendering it uncreative from the user’s perspective. Alternatively, it may be constantly adjusted through human involvement, which impedes the system’s creative autonomy. Finally, a system could be denied self-transformation, which would severely restrict their creative potential. All consequences limit the benefit to their users.

We address these challenges in a particularly difficult scenario: the estimation of a person’s subjective experience of interactive artefacts. We define these as the subset of dynamic, i.e. time-varying, artefacts that are designed to be actively changed by the audience through continuous interaction. Just like the artefacts themselves, their experience through people also varies over time and with respect to an individual’s specific interaction. Conceiving an interactive artefact as a whole would require assessing the experience resulting from
any possible interaction. While we can often formally capture the scope of all possible expressions of interactive artefacts, they practically refuse full experiential access. The estimation of subjective experience of such artefacts is particularly hard, since no designer of a theory-based approach can anticipate all possible expressions, and human reports in a data-driven approach can also only capture part of what we coin the artefacts’ experiential range.

In Sec. 7.3, we propose to mitigate the three shortcomings of existing approaches by estimating the human subjective experience of artefacts via computational intrinsic reward (IR). We tackle the specific challenges of assessing interactive artefacts by simulating people’s interaction with them. We address the time-variance of experience by assessing IR along the whole interaction trajectory. Rather than trying to capture an artefact’s full experiential range, we propose to determine the typical experience of a generic or specific type of audience or an individual by sampling the expected IR for different simulation models. In Sec. 7.4, we instantiate this proposal in a model that uses empowerment (cf. Ch. 3) as the IR.

Many traditionally dynamic artefacts such as installations, dance, music, film and storytelling have been adopted in interactive variants. Videogames in contrast are inherently interactive. In the next section, we assess how our approach can benefit videogame AI as a CC application domain.

### 7.2 Player Experience Modelling on Procedural Game Content

Videogames are the most popular kind of interactive artefact by far; Liapis, Yannakakis and Togelius (2014) note that the ‘play experience is highly interactive and engaging, more so than any other form of art’ (ibid., p. 46). Videogames are a prime candidate for estimating human subjective value as they ‘can be appreciated as an art form (...) only when experienced through play’ (ibid., p. 46, emphasis added). The player experience (PX) describes the personal, transient and dynamic qualities that a player experiences from interacting with a game (Wiemeyer et al., 2016). It expands beyond aesthetics as a sense of beauty and covers a wide range of qualities such as challenge, curiosity, competence, autonomy, relatedness, control, immersion, presence, flow, engagement, tension, and affect (cf. Calleja, 2011).

Videogames are interactive in at least two ways. A game not only changes and unfolds in interaction with a player, but it is also itself a complex product of individual, interacting content facets such as the mechanics, level design, characters, etc. A player’s subjective experience cannot be determined based on the individual facets, but only with respect to the complex whole. This multifacetedness adds to the challenges of evaluating interactive artefacts set out in the previous section. For these and other characteristics, Liapis, Yannakakis and Togelius (2014) have advocated videogames as a ‘killer application’ (ibid., p. 46) for the study of CC. We have chosen games as our application domain for their tough evaluation challenges, but also because they allow to tackle these challenges systematically: their complexity can be kept small to support experimental inquiry, and, as autotelic activities with negotiable consequences (cf. Sec. 5.1.1), we expect PX to be less influenced by external societal and cultural factors that would be hard to control.

### Choice as Application Domain

Videogames are the most popular kind of interactive artefact by far; Liapis, Yannakakis and Togelius (2014) note that the ‘play experience is highly interactive and engaging, more so than any other form of art’ (ibid., p. 46). Videogames are a prime candidate for estimating human subjective value as they ‘can be appreciated as an art form (...) only when experienced through play’ (ibid., p. 46, emphasis added). The player experience (PX) describes the personal, transient and dynamic qualities that a player experiences from interacting with a game (Wiemeyer et al., 2016). It expands beyond aesthetics as a sense of beauty and covers a wide range of qualities such as challenge, curiosity, competence, autonomy, relatedness, control, immersion, presence, flow, engagement, tension, and affect (cf. Calleja, 2011).

Videogames are interactive in at least two ways. A game not only changes and unfolds in interaction with a player, but it is also itself a complex product of individual, interacting content facets such as the mechanics, level design, characters, etc. A player’s subjective experience cannot be determined based on the individual facets, but only with respect to the complex whole. This multifacetedness adds to the challenges of evaluating interactive artefacts set out in the previous section. For these and other characteristics, Liapis, Yannakakis and Togelius (2014) have advocated videogames as a ‘killer application’ (ibid., p. 46) for the study of CC. We have chosen games as our application domain for their tough evaluation challenges, but also because they allow to tackle these challenges systematically: their complexity can be kept small to support experimental inquiry, and, as autotelic activities with negotiable consequences (cf. Sec. 5.1.1), we expect PX to be less influenced by external societal and cultural factors that would be hard to control.
When considering videogames through the lens of CC, researchers usually focus on procedural content generation (PCG) (e.g. Liapis, Yannakakis & Togelius, 2014; Ventura, 2016a). An area of game AI (Yannakakis & Togelius, 2018, pp. 259-260), PCG employs AI techniques ‘for generating game content either autonomously or with only limited human input’ (ibid., p. 151). PCG has a long tradition in games industry, and holds great advantages for modern videogames (Shaker, Togelius & Nelson, 2016, pp. 3-4). By algorithmically creating levels as in Spelunky (Yu, 2008), characters as in Spore (Maxis, 2008), or other elements such as mechanics and music, game developers can satisfy players’ demands for richer and more detailed content, while keeping production costs and time manageable. In addition, players can benefit from content that is automatically tailored to their needs and tastes (Shaker, 2016). Adding new content like quests or weapons procedurally allows designers to create open-ended games, and to increase their replay value (Summerville et al., 2018). PCG relates to CC beyond artefact generation in that it can support designers in becoming more creative: it can encourage new game ideas that only become viable through automation, including games in which PCG constitutes a game mechanic in itself (Cook et al., 2016). Moreover, designers could draw fresh inspiration from the output of PCG algorithms, as they operate under different constraints than their human colleagues.

Designing a PCG system however is tricky: without imposing any constraints, a procedural generator could produce any content instance that its content representation affords, similar to Borges’ (1962) ‘Library of Babel’ containing all possible but mostly meaningless 410-page books. It is a grand challenge of content quality assurance to restrict a generator’s expressive range (Smith & Whitehead, 2010) to those instances that designers and players desire. Crucially, their expectations typically match those of CC researchers towards creative artefacts: adopting the standard definition of creativity (Runco and Jaeger, 2012; and Sec. 4.1.1) and Ritchie’s (2007) empirical criteria for a
computer program to be called creative, the produced artefacts should be *novel*, *typical* and *valuable*. Procedurally generated content is likewise required to be *novel* and *typical*, e.g. a generated quest should be different from existing ones, but still fit the game under consideration. Defining content *value* is more difficult. A generated level should without doubt be *playable*, i.e. there must be a way for the player to succeed or fail. Moreover, they should ideally experience large parts of the content instance rather than just a small fragment. Crucially though, nobody would care about a level, a character or as a consequence even the overall game, if the content in question did not elicit a desired *player experience* (*PX*).

This is particularly striking in games that leverage PCG as their unique selling point. *No Man’s Sky* (Hello Games, 2016) for instance heavily draws on PCG to create a vast number of planets for players to explore. But while different planets appeared varied at first, ongoing exploration could not satisfy players’ curiosity: the *expressive range* of worlds was small after all (Fig. 7.1). The game has consequently been described as ‘infinitely boring’ (Martin, 2016). We hold that *PX* substantially determines the *value* of content instances, and consequently the acceptance and replayability of games (cf. Smith, 2014b). A key challenge in designing a PCG system is hence to realise *experiential control*, i.e. the capacity to ‘control for the kind of experience the player will receive’ (ibid., p. 921). In this chapter, we contribute a new approach to predicting *PX* as a major determinant of game content value to serve the experiential control of procedural generators.

To motivate our contribution, we first highlight the specific challenges of assessing *PX* in PCG by contrasting this task with the non-computational evaluation of *PX* in *games user research*. Here, *PX* is assessed e.g. to give directions at key points of game development, or to infer more general insights about how people interact with games. Assessment is usually done on very few conditions, and during or after their experience through the player. To make detailed and accurate judgements, researchers rely on rich subjective player feedback gathered from e.g. questionnaires or interviews (Boyle et al., 2012; Cairns, Cox & Nordin, 2014; Wiemeyer et al., 2016) and objective player data e.g. in the form of behavioural or physiological observations (Nacke, 2013). However, these methods are typically expensive, slow and difficult to perform. Participants want to be paid, and the evaluation speed is limited by their game-playing ability and information processing capacity. Moreover, subjective measures become easily biased, and objective measures are sensitive to environmental factors, hence necessitating strong experimental control.

These methods can still be employed when PCG is used as a tool by game designers during development, providing content that is then carefully curated before inclusion into the game (cf. Craveirinha, Barreto & Roque, 2016). However, the full power of PCG, especially for replayability and customisation, can only be unleashed when it is used in the shipped game, either offline before, or *online* during play. In this scenario, usually many content instances must be assessed before they are experienced by a player. Each instance can be considered one experimental condition resulting from e.g. different parameter combinations in the generator. Speed is of prime importance, and the subjective and objective approaches of games user research do not scale.
To elicit a particular PX nonetheless, designers employ constructive algorithms that run in fixed time and do not involve any explicit evaluation. The quality of content is assured by only performing operations ‘that are guaranteed to never produce broken content’ (Togelius et al., 2011, p. 174). A popular constructive technique is to chain together experiential chunks (Smith, 2014b), i.e. content fragments that have been manually assessed for a certain experience during development. Fig. 7.2 illustrates how level segments as experiential chunks are procedurally arranged for the Underground expansion of Tom Clancy’s The Division (Massive Entertainment / Red Storm Entertainment, 2016). This approach is used to retain experiential control in many other games such as Spelunky (Yu, 2008) and Diablo 3 (Blizzard Entertainment, 2012). These and related constructive approaches such as the use of generative grammars in No Man’s Sky (Hello Games, 2016) make the explicit (Craveirinha, Barreto & Roque, 2016) prediction of PX obsolete, but they considerably limit the expressive range of generators.

Researchers have sought to overcome the limitations of these constructive approaches by leveraging models of PX in PCG as a means to automatically evaluate candidate content. Yannakakis and Togelius (ibid.) classify existing work in their experience-driven PCG framework. Their taxonomy distinguishes components of the (i) content evaluation, encompassing how PX is modelled and how it is assessed on content instances, from (ii) generative components, capturing how content is represented and generated. In this chapter, we present a complete experience-driven PCG approach, but our core innovation concerns the modelling of PX (i). We motivate our contribution by addressing a shortcoming of present experience-driven PCG. To this end, we first distinguish how existing approaches to PX modelling involve people in development.

Yannakakis and Togelius (ibid.) distinguish three approaches to designing a PX model. The first two leverage subjective or objective player data as model input features or experience labels. For data acquisition, these approaches rely on the same methods as games user research. The third, gameplay-based approach uses ‘statistical spatio-temporal features of game interaction’ (ibid., p. 153) acquired from either human or simulated play. It also encompasses static game features that would shape this interaction, e.g. properties of a game’s level structure. For the sake of our argument, we distinguish the sub-
categories of model-free and model-based gameplay PX models. In the first case, the mapping between gameplay features and PX is automatically learned from player data. In model-based approaches in contrast, the mapping is manually established by the model designer, informed by intuition or theoretical frameworks. Both approaches can be combined in hybrid models.

Most existing work leverages multiple approaches, and we provide four examples to illustrate the above differences. Yannakakis, Martínez and Jhala (2010) for instance use subjective preference ratings jointly with objective data to learn linear and non-linear PX models on a 3D predator-and-prey game. They assess players’ heart rate, pulse and skin conductance to predict a wide range of experiences such as fun, challenge, boredom, etc. reported through a preference rating questionnaire. Guzdial, Sturtevant and Li (2016) also employ a model-free approach, but combine both gameplay features and subjective player data to estimate players’ enjoyment, difficulty and visual aesthetics of Infinite Mario Bros. (Persson, 2010) levels. They use player ratings on a large corpus of levels from an existing dataset (Reis, Lelis et al., 2015) to train a convolutional neural network to automatically extract content features that correlate with these experiences. Sorenson, Pasquier and DiPaola (2011) focus on the same game, but choose a model-based approach to determine players’ enjoyment of levels based on gameplay features such as the size of gaps, the player’s maximum jump length and the presence of enemies. Their hand-crafted model is informed by several theories of fun (e.g. Koster, 2013), flow (Sweetser & Wyeth, 2005) and challenge (Salen & Zimmerman, 2004) and estimates players’ enjoyment of levels based on the presence of different ‘rhythm groups’, i.e. ‘alternating periods of high and low challenge’ (Sorenson, Pasquier & DiPaola, 2011, p. 243). While the earlier examples rely on static gameplay features, Togelius, De Nardi and Lucas (2007) evaluate racing game tracks via gameplay data from simulated play. They first train an AI agent on human play data, and then use it to estimate a player’s level progress, performance variation and difference in driving speed on procedurally generated tracks. Their model-based approach draws on Malone’s (1980) heuristics for engaging games and Koster’s (2013) theory of fun to map this data to player enjoyment.

Acquiring subjective and objective player data during model development suffers from the same drawbacks as discussed earlier in the context of games user research. Alternatively, such data can be obtained online through interactive approaches to PX modelling (Yannakakis & Togelius, 2011), but this is mostly practically infeasible: while it is possible to measure certain subjective experiential preferences implicitly (e.g. Hastings, Guha & Stanley, 2009), richer measurements require explicit interactions, e.g. through questionnaires, which is usually too obtrusive. The same applies for the assessment of objective player data, which is moreover mostly technically implausible in commercial games outside a lab environment. Yannakakis and Togelius (2011) point out that gameplay-, model-based approaches constitute the least intrusive option,

---

4 Yannakakis and Togelius’ (2011) distinction between model-free and model-based approaches may seem confusing, as the first type of approach typically rests on machine-learning models to learn the relationship between input features and PX. We adopt their terminology for comparability, but promote the alternative notions of data-based and theory-based approaches.
but usually have a lower resolution as they rest on strong assumptions on the relationship of gameplay and PX.

These drawbacks are well known. We motivate our contribution based on a separate shortcoming that is common to all approaches but has yet received little attention: present PX models are too dependent on continuous human involvement. In particular, existing models always rely on people in their development\(^5\), and we argue that this renders them inflexible: they likely become inaccurate as soon as new elements are introduced to a game, present ones are altered, or parameters of the procedural generator are changed. This is because games are multifaceted, complex systems: the PX emerges from the interplay of many different content facets. Adding the Berserk power-up to a battle scene in *Doom* (id Software, 1993) for instance is likely to change PX drastically: if collected, the player’s health is restored and their weapon damage multiplied, making the same scene considerably less challenging. Once fixed, existing PX models are usually too specific to maintain their accuracy in light of such changes. Depending on the underlying approach, they would need to be either retrained based on new feedback from players, or manually re-adjusted by their designers. We deem it infeasible for designers to manually craft more flexible models, as they can anticipate the possible effects of content facet interactions on PX only to a minor extent. Crucially though, such changes to the game and content generator are commonplace during development, even in the late phases when e.g. a game is being balanced. Existing approaches hence not only rely on the involvement of players and designers at some point, but continuously. The associated drawbacks outlined earlier, including the time demand and costs, are hence recurring.

These shortcomings align with the context-specificity of existing approaches to evaluating subjective experience in CC, and their dependency on people (cf. Sec. 7.1). Addressing these challenges in videogame AI could hence yield insights for CC more generally. PX substantially contributes to a game’s value, which in turn is essential for determining the creativity of these multifaceted, interactive artefacts. Liapis, Yannakakis and Togelius (2014) note that evaluating ‘compound game creativity which treats the game as a coherent entity and not the sum of its parts is a key research question which can potentially lead to breakthroughs in creativity research’ (ibid.).

Within the application domain of game AI, our immediate goal is to develop PX models that are more robust to changes in a specific game or content generator. Our long-term vision is for models to become game-general (Togelius & Yannakakis, 2016), and hence suitable for application in automated game design (e.g. Cook, Colton & Gow, 2016a, 2016b). To succeed, we must address a somewhat paradoxical challenge: can we overcome the continuous involvement of players and designers in the design of models that predict human experience? Can we develop models that remain accurate across content modifications without involving people again, thus facilitating fast and cheap content evaluation?

\(^5\) We ignore interactive approaches here for the drawbacks elaborated earlier. Acquiring explicit, subjective player feedback is rarely an option, as the process is either too obtrusive, e.g. when using questionnaires, or too imprecise, e.g. when asking for binary preference ratings. Assessing objective player feedback interactively is usually technically implausible.
Our work is in the tradition of Nelson (2011), who has set out to evaluating games *without empirical player data*. While he has proposed strategies to uncovering which states a specific game can realise, and to explore the boundaries of that possibility space, our work pushes this agenda further towards modelling PX. In the next section, we introduce our novel approach to modelling PX without continuous player or designer involvement.

### 7.3 Predicting Player Experience via Intrinsic Reward

In an effort to overcome the shortcomings of existing PX models in experience-driven PCG, we propose a more *flexible* alternative to modelling a player’s experience that is *independent* of human players, and does not rely on designers’ knowledge about a game’s semantics. Our approach models PX based on computational IR calculated on the simulated gameplay of AI agents. In this section, we motivate and introduce this approach informally and independently of a specific reward formalism. In Sec. 7.4, we then instantiate it based on empowerment as intrinsic reward (Ch. 3).

Following Yannakakis and Togelius’ (2011) taxonomy, we distinguish two evaluation components: the **PX model** and the **content quality assessment** method. Roughly speaking, the first defines the type of data used, the modelled experiences, and the mapping in-between. The second specifies how this data is acquired and how the measurement is accomplished.

Our model estimates human PX through *computational IR*, calculated on *gameplay features* in the form of a play trajectory through the *game state* space. We consider IR a natural fit to predict PX for several reasons. IM has been closely linked to human gameplay in game design theory (Sec. 5.1.1), and to PX in games user research (Sec. 5.1.2). By definition, IR is independent of a specific instrumental outcome (Sec. 2.1 and 2.2.2), and we thus expect it to remain a valid, generic predictor of PX within and across different games, making continuous re-adjustments through the model designer unnecessary. These definitional and empirical arguments are complemented by the formal properties of IR (Sec. 2.2.3). Being *agent-centric*, IR can be calculated independently of any game-specific, external goals. Its *embodiment universality* allows IR to be calculated across, and yet to be sensitive to the different ways that a player avatar can interface with the game world. Being *free of semantics*, the calculation of IR is independent of designers’ knowledge of the meaning of game tokens, potentially enabling its application across different games without adjustments. Jointly, these properties imply that IR is not specific to one but sensitive to the interaction of many content facets, thus addressing the challenge of quantifying the experience of games as *multifaceted* artefacts. In summary, IR is a promising candidate for predicting PX independently of human players and designers across games and game facets.

We calculate IR as PX predictor in all states along an assumed gameplay trajectory. For our content quality assessment to be as independent of people

---

6 There is a strong overlap between these and the reasons to embrace IM for NPCs and player modelling, identified in our systematic review of existing game AI work in Sec. 5.2.1.

7 We focus on our videogame AI application here, but note that its *agent-centricity* and *embodiment sensitivity* allows for IR to quantify subjective differences, and its use hence accommodates the critique of CC evaluation methods to be *subject-unspecific* (Sec. 7.4).
as possible, we sample such trajectories via the simulation of AI game-playing agents. We adopt this approach from existing work as it allows for the unobtrusive and efficient evaluation of content candidates. Combining it with IR moreover allows us in principle to assess PX dynamically over the course of game-play. We hence consider it a potential non-obtrusive alternative to measuring objective, time-variant player data.

We complement this componential account with a description of the evaluation procedure for a given content candidate. We assume to be given a (partial) game as input that features the content candidate to be investigated. The calculation of a PX prediction then comprises three steps:

1. **Recording state traces** Perform one or multiple simulations of AI gameplay on the input game. Each simulation is executed until the agent ends up in a terminal state or hits a time limit. For each gameplay instance, record the game states encountered in order. The goal is to identify the typical states that human players would encounter when playing the given game. Implicitly, this also serves the identification of inaccessible states that are theoretically permitted by the game state definition, but never realised at runtime. The AI agent’s policy can be chosen to optimise the same reward used for PX prediction (on-policy), to follow different goals (off-policy\(^8\)), or to realise any combination of the two. It could be provided by a general game-playing agent, be parametrised to reflect certain player types, or be adapted during runtime to fit the style of a particular human player. Each option has implications on the generality of the overall approach.

2. **Calculating intrinsic reward** Intrinsic reward is calculated for each state in the individual recorded gameplay traces. For adaptive (cf. Oudeyer and Kaplan, 2007; and Sec. 2.2.3) intrinsic reward functions, this calculation must respect the original order of the game states. Static intrinsic rewards afford the more efficient procedure to evaluate identical game states only once, and weight the rewards accordingly.

3. **Aggregating rewards into prediction** The intrinsic rewards are aggregated into a scalar experience prediction. This transformation does not have to, but can be sensitive to the order of intrinsic reward\(^9\). If multiple state traces are available, it should include an expectation over all available samples, potentially taking into account the likelihood of each sample to reflect human gameplay. Normalisation is required to allow for the comparison of predictions on different content instances.

If the IR function is adaptive, it is likely more efficient to interleave step (1) and (2), i.e. to simulate gameplay and calculate IR within the same time step before commencing the simulation.

---

\(^8\) Similar to Sec. 6.4, we use the off-policy notion to highlight a discrepancy between an assumed and an executed policy: an experience estimate would be calculated based on the assumption that the player is at least partly acting to optimise a specific intrinsic reward, but at the same time, this reward is calculated along gameplay simulations driven by a different policy.

\(^9\) The prediction could e.g. be given by the fit of the intrinsic reward sequence to an experience curve, similar to the ‘rhythm groups’ investigated by Sorenson, Pasquier and DiPaola (2011).
We hypothesise that different IR functions can be used to predict different PXs. The choice of a specific reward function should be informed by insights in game design theory, games user research, and other relevant human-computer interaction (HCI) research. Any such instantiation essentially represents our, i.e. the researcher’s or the game designer’s, theory of mind (Premack & Woodruff, 1978) with respect to the human player. Not only do we assume the human player to follow a certain policy, we also expect intrinsic reward to affect both cognitive and affective states, and thus cross two camps of established theories (Melhart, Yannakakis & Liapis, 2020).

Within the experience-driven PCG framework Yannakakis and Togelius (2011), our approach represents a combination of gameplay- and model-based PX modelling and simulation-based content quality assessment. It has been suggested that model-based approaches come with a lower resolution than those using subjective and objective player data (Yannakakis and Togelius, 2011; and Sec. 7.2); we do not primarily target resolution, but aim to improve generality.

7.4 Empowerment-Based Player Experience Prediction

We instantiate our proposed approach in an empowerment-based player experience prediction (EBPXP) model. We first defend our choice of empowerment (Ch. 3) as IR based on its potential role in human gameplay, assessed by its proximity to game design and games user research theories, and preliminary empirical findings on its relationship to human experience beyond games. We then formalise our approach and illustrate its calculation via pseudocode.

7.4.1 Choice of Empowerment

The maximisation of empowerment as an agent’s perceivable influence has been hypothesised as a unifying principle explaining many behavioural phenomena throughout the animal kingdom (Sec. 3.1). Some argue that ‘understanding games is approaching a phase where it is close to understanding the psychology of individual life experiences in general’ (Takatalo et al., 2010, p. 25); games borrow from the mechanics that govern our reality, and it is hence little surprising that EM has been successfully used in driving the behaviour of general game-playing agents (Sec. 5.2.1) and NPCs (Ch. 6) towards realising a game’s latent goals. Games are designed to be intrinsically motivating for people (Sec. 5.1.1), and we have illustrated in Sec. 6.4 that many instances of human gameplay can be modelled as empowerment maximising decision-making processes. In deciding on an action, IR distinguishes preferable from less preferable situations, and it likely also shapes our fine-grained experiences of these situations. In summary, we hypothesise that empowerment influences how people play and experience games.

Our hypothesis is supported by several theories in game design. Salen and Zimmerman (2004) describe play in the most general sense as ‘free movement within a more rigid structure’ (Salen and Zimmerman, 2004, p. 304; and Sec. 5.1). As a measure of controllability and observability (Sec. 3.2), empowerment could allow players to trace these boundaries in play. For the same reasons,
7.4 Empowerment-based Player Experience Prediction

Empowerment may correlate with the experience of *tegotae*, described by Shigeru Miyamoto as the satisfaction from being in control (Perry, 2016). The creator of *Super Mario Bros.* (Nintendo R&D, 1985) puts it at the core of his games design: ‘if the player jumps from a high place, the amount of time they stay in the air needs to be just right, or else they’ll feel disconnected from the experience’ (Perry, 2016). The player’s empowerment would be zero while being in the air. Empowerment moreover captures various factors of *outcome uncertainty*, which Salen and Zimmerman (2004) deem essential for meaningful play: ‘If the outcome of the game is completely predetermined – then any choices a player makes are meaningless, because they do not impact the way that the game plays out’ (ibid., p. 174). Caillois (1961) directly relates outcome uncertainty to *enjoyment*. Empowerment measures the consequences of a player’s action sequences, and could, given a sufficiently large lookahead, quantify the breadth of a game’s outcomes to the extent that they can be perceptually distinguished by the player. More commonly though, empowerment captures the richness of intermediate game states, as the lookahead is computationally restricted. Empowerment moreover quantifies the controllability of action outcomes as a second factor of outcome uncertainty: it decreases the more the outcomes of different actions overlap (Sec. 3.2).

We suggest further links between this IR and experiences investigated in games user research. We hypothesise empowerment to be closely related to a player’s *effectance* (Sec. 5.1.2) as the experience of ‘receiving immediate, direct feedback on one’s action and of influencing the game world’ (Klimmt, Hartmann & Frey, 2007, p. 845). More coarsely, we consider its relationship to two components of *self-determination theory* (Ryan & Deci, 2000b) (cf. Sec. 3.1). Empowerment relates to in-game *autonomy* as ‘the degree of choice one has over the sequence of actions, or the tasks and goals undertaken’ (Ryan, Rigby & Przybylski, 2006, p. 349) in that it quantifies the availability of actions in different game states. Moreover, its proximity to a player’s effectance also suggests a close connection to their perceived *competence* as the ‘need for challenge and feelings of effectance’ (ibid., p. 349).

Preliminary empirical support for the relationship of empowerment and human experience more generally comes from HCI research, where Trendafilov and Murray-Smith (2013) have used empowerment to quantify the impact of uncertainty on human experience in manual control. Interfaces with reduced empowerment were correlated with feelings of *frustration*. Crucially though, no direct connection between empowerment and player experience has been manifested yet, motivating our study in Sec. 7.5. Based on these leads, we choose empowerment as the foundation for our PX prediction approach.

7.4.2 Formalisation

To highlight how EBPXP instantiates our proposal of predicting PX via IR, we formalise it along the same structure used in Sec. 7.3: we cover the player experience model, the content quality assessment method and finish with a description of the evaluation procedure supported by pseudocode. First of all though, we lay out our formal assumptions.
Our formalisation of EBPXP and the calculation of player empowerment rests on the framework developed in Sec. 3.2. We model the interaction of a single player with the game by means of the memoryless perception-action (PA)-loop in Fig. D.1b. In this causal Bayesian network, the **game state** at time \( t \) is represented by the random variable \( R_t \), and the **player** by their sensor perceptions \( S_t \) and actions \( A_t \). A player may only have limited insight into a game’s actual workings, and we hence distinguish between the game’s **objective dynamics** \( p(\ldots) \), and the player’s **models** of these dynamics \( q(\ldots) \).

The EBPXP model at present draws on a simplified account of the player-game interaction. We firstly assume **full observability**, i.e. the player can perceive the whole game state. We thus have \( s_t = r_t \) and can express the sensor and environment dynamics in a single sensorimotor distribution. We secondly assume **fixed parameters**: we can imagine that the player already had a chance to familiarise themselves with the game and arrived at a stable estimate of their sensorimotor dynamics and policy. We do not infer the corresponding parameters, and hence omit them from the distributions. Based on these simplifying assumptions, we have the objective and modelled sensorimotor dynamics \( p(s_{t+1}|a_t, s_t), q(s_{t+1}|\hat{a}_t, \hat{s}_t) \) and the objective policy \( p(\hat{a}_t|\hat{s}_t) \). We denote variables that are assumed by the player but have not yet been resolved or realised by a hat, and intervened (Appx. A) variables by a dot. We thus use the same assumptions as in Sec. 3.2 to formalise simplified versions of various empowerment quantities, and draw on the respective earlier equations.

Our **player experience model** as the first evaluation component comes down to the IR function used to assess game states. For EBPXP, we want to estimate a player’s experience in **decision-making**, i.e. at a point where they can still compare action alternatives based on the goodness of the situations which these actions are expected to yield. Rather than using a player’s vanilla empowerment in the present situation, we hence assess PX based on their **state-expected empowerment** directed one time step into the future. It corresponds to the expected empowerment over all successor states that could be caused by all available actions in the present state:

\[
E(s_{t+1}, \hat{A}_t|\hat{s}_t) = \sum_{\hat{a}_{t+1}} q(\hat{s}_{t+1}|\hat{a}_t, \hat{s}_t) q_U(\hat{a}_t|\hat{s}_t) E(\hat{s}_{t+1})
\]  

(7.1)

Here, \( q(\hat{s}_{t+1}|\hat{a}_t^n, \hat{s}_t) \) are the 1-step sensorimotor dynamics, \( q_U(\hat{a}_t|\hat{s}_t) \sim U(|\hat{A}_t|) \) is the uniform policy, and \( E(\hat{s}_{t+1}) \) is the simplified vanilla empowerment at the next time step, as originally defined in Eq. 3.14:

\[
E(\hat{s}_t) = \max_{q(\hat{a}_t^n)} I(\hat{A}_t^n \rightarrow \hat{s}_{t+1}|\hat{s}_t)
\]

\[
= \max_{q(\hat{a}_t^n)} \sum_{\hat{a}_t^n, \hat{s}_{t+1}} q(\hat{a}_t^n) q(\hat{s}_{t+1}|\hat{a}_t^n, \hat{s}_t) \log \frac{q(\hat{s}_{t+1}|\hat{a}_t^n, \hat{s}_t)}{\sum_{\hat{a}_t^n} q(\hat{s}_{t+1}|\hat{a}_t^n, \hat{s}_t) q(\hat{a}_t^n)}
\]

The term \( q(\hat{s}_{t+1}|\hat{a}_t^n, \hat{s}_t) \) represents the recursively calculated, simplified n-step sensorimotor dynamics (Eq. 3.16). We use the uniform policy in the expectation (Eq. 7.1) to express an assumption of maximum ignorance about
the next action to be chosen at the time of experience assessment. Similarly, vanilla empowerment is calculated in an open-loop fashion (Sec. 3.3).

To predict the PX of a content instance, we must consider its interaction with other facets of the game in which the content is embedded. We hence perform quality assessment on the whole game and a specific content instance. Hence, we use the PX prediction method to predict the PX of a content instance, which is defined as follows:

\[ \text{Predicted PX} = \mathbb{E}[\mathcal{E}(\tau)] \]

where \( \mathcal{E}(\tau) \) is the expected empowerment reward for a sequence of actions \( \tau = (s_0, a_1, s_1, \ldots, s_{\tau-1}) \).

Algorithm 2 Empowerment-based player experience prediction (EBPXP)

1: \textbf{function} EBPXP(\( g, p(a_t|s_t), n, q(s_t+1|\hat{a}_t, s_t), T \))
2:  
3: \hspace{1em} \text{Stage 1: Recording state trace}
4: \hspace{2em} Simulate AI play of game \( g = (s_0, p(s_{t+1}|a_t, s_t), S_T) \) with policy \( p(a_t|s_t) \) and record state trajectory \( \tau \) until terminal state hit
5: \hspace{2em} \( t \leftarrow 0 \), \( \tau(0) \leftarrow s_0 \)
6: \hspace{2em} \( s_t \leftarrow \tau(t) \)
7: \hspace{2em} while \( s_t \notin S_T \) and \( t < T \) do
8: \hspace{3em} \( a_t \sim p(a_t|s_t) \)
9: \hspace{3em} \( s_{t+1} \sim p(s_{t+1}|a_t, s_t) \)
10: \hspace{3em} \( \tau(t+1) \leftarrow s_{t+1} \)
11: \hspace{3em} \( t \leftarrow t+1 \)
12: \hspace{2em} end while
13: \hspace{2em} for \( \hat{a}_t \in \hat{A}_t \) do
14: \hspace{3em} \( \text{for } \hat{s}_{t+1} \in \hat{S}_{t+1} : (q(\hat{s}_{t+1}|\hat{a}_t, s_t) > 0 \land \mathcal{C}(\hat{s}_{t+1}) = \emptyset) \) do
15: \hspace{4em} Calculate \( q(\hat{s}_{t+n+1}|\hat{a}_t^n, \hat{s}_{t+1}) \) recursively (Eq. 3.16)
16: \hspace{4em} Find \( q^*(\hat{a}_t^n) \) that maximises the channel capacity for \( q(\hat{s}_{t+n+1}|\hat{a}_t^n, \hat{s}_{t+1}) \) with the Blahut-Arimoto algorithm
17: \hspace{4em} \( \mathcal{E}(\hat{s}_{t+1}) \leftarrow I(\hat{A}_{t+1} \cup \hat{S}_{t+1}) \rightarrow \hat{S}_{t+n+1} \) for \( q^*(\hat{a}_t^n) \) (Eq. 3.14)
18: \hspace{3em} end for
19: \hspace{2em} end for
20: \hspace{2em} \( \mathbb{E}[\mathcal{E}(\tau)] \leftarrow \sum_{\hat{a}_{t+1}, \hat{s}_{t+1}} q(\hat{s}_{t+1}|\hat{a}_t, s_t)q_f(\hat{a}_t|s_t)\mathcal{E}(\hat{s}_{t+1}) \) (Eq. 7.1)
21: \hspace{2em} end for
22: \hspace{2em} end for
23: \hspace{1em} end function
acknowledge that a player’s behaviour may not be driven by EM alone, but yet assume that empowerment can capture a part of their experience.

Our PX prediction, and hence the output of EBX, is given by the mean state-expected empowerment over the whole trajectory:

\[
\mathbb{E}[\mathcal{E}](\tau) := \frac{1}{|\tau|} \sum_{t=0}^{|\tau|-1} \mathbb{E}_{S_{t+1}, A_t | s_t}[\mathcal{E}]
\]

At this point, we do not consider the order of rewards in our prediction. The rationale behind this is to not limit our approach with further theoretical assumptions about the relationship of dynamic reward structures and PX.

We describe EBX for a given content candidate via the pseudocode in Alg. 2. The evaluation procedure follows the same three steps as the generic approach outlined in Sec. 7.3. As a consequence of our simplifying assumptions, our game-playing agent as well as their IR are static (Oudeyer & Kaplan, 2007; Yannakakis & Togelius, 2011), i.e. neither the agent nor their reward change over time. This allows us to speed up the computation slightly: for recurring game states (line 13), we only have to calculate their state-dependent n-step empowerment (line 16) once. Amongst others, the algorithm is parametrised by \( T \), the maximum length of a trajectory, ensuring that it does not loop indefinitely for a playing agent that gets stuck in non-goal states.

EBX comes with only few requirements: Firstly, the calculation of empowerment relies on access to a model of the game’s dynamics. This does not have to be the actual game state forward model, which is often unavailable or might be computationally expensive to evaluate. Empowerment quantifies an agent’s perceivable control, and even a perfect dynamics model thus does not need to reflect all action-induced changes to the game’s global state, but only local changes that can be perceived by the player. One could thus refrain to a local, more efficient model variant. Moreover, one can use an imperfect, simplified forward model, potentially learned from the simulated player’s experience, for the calculation of epistemic empowerment (cf. Sec. 3.2).

Secondly, EBX requires an AI game-playing agent’s policy to sample game state trajectories. This policy can be highly general, or describe a group or even a specific player’s behaviour. The choice of model likely has implications on EBX’s prediction accuracy and its generality, and we discuss this trade-off in Sec. 7.6. The agent’s sensorimotor dynamics model is an optional requirement which is only needed if epistemic uncertainty (Appx. A) on the agent side is to be captured, or if the objective dynamics, i.e. the game’s forward model, is unavailable or must be simplified.

7.5 EXPLORATORY STUDY

We propose to predict human PX based on IR calculated along AI-simulated play trajectories. Our goal is to unleash the procedural generation of game content and entire games as the most popular part of computational game creativity (Liapis, Yannakakis & Togelius, 2014) and a CC application domain. Through this use-case, we more generally aim to contribute to modelling people’s subjective experiences of interactive artefacts. Our study represents...
the first step towards a proof-of-concept based on EBPXP, which instantiates our generic approach in Sec. 7.3. Our findings contribute to the overarching research question of this chapter: ‘Can we use IR to predict people’s experience of interactive artefacts in a general and autonomous way?’ (RQ.9).

The formal interpretation of empowerment as an AI agent’s potential and perceivable influence over the environment (Sec. 3.2) does not warrant any conclusions on how the computational IR relates to human PX. In order to conduct a quantitative study based on established, reliable HCI instruments for a full proof-of-concept in future work, we must first identify candidate experiences which empowerment could potentially predict.

We focus on predicting PX for procedurally generated levels of an infinite runner game specifically developed for this study. Levels are the arguably most popular game facet in PCG research, and procedural level design represents ‘one of the oldest and most popular commercial applications of autonomous creative systems’ (Liapis, Yannakakis & Togelius, 2014, p. 49). Moreover, levels substantially contribute to a game’s identity: ‘A game’s tone is often set by its levels and the challenges they pose; digital games often have a constant or near-constant set of mechanics throughout, but vary the gameplay and challenge through level design’ (ibid., p. 49). Our study probes the prediction:

PD.6 Game levels with different mean state-expected empowerment are experienced differently by human players.

We not only investigate whether such levels are experienced differently, but also explore which specific PXs our approach could predict. To this end, we conduct a qualitative study of human player think-alouds based on experiential vignettes (Hudson & Cairns, 2014a), and assessed through a thematic analysis (Braun & Clarke, 2006). We describe our study details and how the conditions have been generated, and then report and discuss our results. We eventually inform a hypothesis for a future quantitative study by relating our results back to game design and games user research.

7.5.1 Methodology

We approach this study as an instance of an experiential vignette (Hudson & Cairns, 2014a). This qualitative method has been developed to investigate user experience phenomena in digital games that have ‘not yet been well defined and understood’ (ibid., p. 103), and are not amenable to quantitative measurement as this would constrain the richness and complexity of responses. Experiential vignettes are different from other qualitative approaches in that the participant is exposed to well-defined situations that aim to ‘manipulate’ their responses in order to probe certain aspects of the phenomenon. To probe PD.6, we expose our participants to different levels of an (in)finite runner game. Each level has been selected for having a different mean state-expected empowerment. These conditions are predicted to manipulate the player’s experiences, and the experiment thus qualifies as an experiential vignette. To gather unbiased qualitative data, we ask participants to think aloud while playing the game. We complement this with few targeted questions.

We perform a thematic analysis (Braun & Clarke, 2006) on this qualitative data to find out which types of experiences our conditions jointly give
rise to, and how these experiences differ between them. A thematic analysis is a qualitative method ‘for identifying, analysing and reporting patterns (themes)’ (Braun & Clarke, 2006, p. 79) across individual items of a data set, in our case across individual recordings of player think-alouds. We have chosen this method as it results in a ‘thick description’ (ibid., p. 97) of the dataset which can uncover unanticipated insights to benefit our exploration. At the same time, a thematic analysis is less complex than other qualitative methods such as grounded theory (Salisbury & Cole, 2016). We decided against a more quantifiable approach such as content analysis (Mayring, 2004), since our goal was not to see how often people engage in a set of previously known experiences, but to explore the range of yet unknown experiences in response to our manipulation. Yet, the prevalence of specific experiences in the dataset provides us with clues about which PXs our approach could predict best. Given the ambiguity, diversity and differing foci of participants’ utterances in the think-aloud data, we however only coarsely state such prevalence.

We adopt Braun and Clarke’s (2006) taxonomy to clarify the specific flavour of our thematic analysis. We conceive a theme as a specific PX that has been coded in at least two think-alouds. Our analysis is more inductive than theoretical, in that our identified themes are closely linked to the data itself, rather than to any specific questions asked in the study or to the experimenter’s preconceptions. We identify our themes mostly at a latent rather than a semantic level; instead of only considering the ‘explicit or surface meanings’ (ibid., p. 84) of the player’s utterances, we try to identify the features that give rise to this meaning. More specifically, we identify these utterances with known PXs, and seek to explicate their emergence, e.g. from a level’s structure, based on existing theoretical frameworks. Moreover, our method is realist rather than constructivist in that we report the ‘experiences, meanings and the reality of participants’ (ibid., p. 81), assuming that their ‘language reflects and enables us to articulate meaning and experience’ (ibid., p. 85).

We have evaluated player’s responses to levels with low and high mean state-expected empowerment. We refer to these conditions as low and high, respectively. The levels were procedurally generated, and, due to stochasticity in the generator, vary locally despite having the same mean state-expected empowerment. We have chosen a mixed experimental design to balance effects of these within-condition differences on PX. In our experiment, the two conditions are evaluated within subjects, while two instances of the same condition are considered between subjects. Each participant first had to complete one of two tutorial levels with medium state-expected empowerment to familiarise themselves with the game. We thus presented our participants with eight level combinations (2 instances × 2 conditions × 2 tutorials) in balanced order. Since the resulting data is qualitative and analysed extensively, our experiential vignette only requires a modest number of participants.

7.5.2 Testbed

Participants were asked to play different levels of RoboRunner (Fig. 7.3), a one-button game from the infinite runner genre specifically developed for this study. The goal is to drive a yellow robot from the left to the right through a
space station to escape imminent self-destruction. While RoboRunner realises most characteristics of infinite runner games, our levels are finite to bound the variation in shared game content that participants with different skill can experience. In order to reach the end of the level and thus be spared, the robot has to jump across chasms. The game has one implicit (do nothing) and one explicit action: the robot drives to the right automatically with constant speed, and the player can invoke jumps with a button press when their avatar is on the floor. RoboRunner is completely deterministic to guarantee that different runs of the same condition in principle yield the same gameplay.

We have chosen this particular testbed for several reasons. Infinite runners are, despite their simplicity, very popular, especially on mobile platforms and with casual gamers. Our participants have likely seen or played similar games such as Canabalt (Saltsman, 2009) before, which should have made it easier for them to familiarise with our testbed. Due to its linear nature, RoboRunner allows us to determine the relevant states the player is going to pass through with little ambiguity: while they can jump at different positions, there are no alternative routes to reach the end of the level; the player can only briefly branch off but eventually returns to a common trajectory. The simple controls and linearity also eased the task to create an AI controller which resembles human play. To the player, the game appears continuous in space, but it is composed of modular blocks and can consequently be treated as discrete to simplify simulated game-playing, the calculation of IR and our analysis of the interview data. The player is the only dynamic element and each game state is fully characterised by their position. Since the game is also deterministic, the state space remains small and tractable for the exhaustive computation of IR. Furthermore, the game’s low complexity allows us to develop a level generator that can sample from the entire expressive range of possible level representations, and thus introduces no bias prior to our selection of experimental conditions.

7.5.3 Materials

To produce the conditions for our study while avoiding a design bias, we have employed search-based PCG (Togelius et al., 2011) based on a genetic algorithm (Eiben & Smith, 2015). Fig. 7.4 shows the user interface of the custom level generator. We use a 1:1 genotype-phenotype mapping: the platform layout is represented as a bit array, with 1 indicating the presence of a platform module and 0 its absence. The phenotypes are rendered two ways: as a design visualisation in the generator interface (Fig. 7.4 bottom), and as
7.5 Exploratory Study

Figure 7.4: Interface of the procedural content generator for RoboRunner levels. The lower part comprises a design visualisation of the previously evolved level. It shows an AI playtrace, empowerment values at individual positions and mean empowerment values per platform. The parameters for the empowerment calculation and evolution of new level instances are set in the top-left window. The middle window allows to import and export levels and control the visualisation. The right-hand window show a log with details about the evolutionary process, and about the empowerment calculation at the level position highlighted in red.

The actual level to be played by the participant (Fig. 7.3). Fig. 7.5 illustrates the genotype-phenotype mapping for the latter. The first five and last five modules of a level are assumed protected to form a ‘safety zone’ for the player.

The loss calculation for a specific level rests on its evaluation through EBPXP as described in Sec. 7.4. We use the following inputs to the model:

- The RoboRunner game $g = (s_0, p(s_{t+1}|q_t, s_t), S_T)$. We assume that each level to be evaluated is wrapped into a separate game instance, which is otherwise identical. The game’s initial state $s_0$ is given by the agent’s position on the left-hand side of the level. The set of terminal states $S_T$ is given by all possible positions on the very right-hand side of

Figure 7.5: The RoboRunner 1:1 genotype-phenotype mapping from a bit array to the platform layout. We show a slice from the beginning of an example level. 1’s in the bit array correspond to platforms and 0’s to gaps in the floor.
a level, mid-air or on a platform. The forward model \( p(s_{t+1}|a_t, s_t) \), is
deterministic, and realises the game’s previously introduced mechanics.

- The AI player policy \( p(a|s_t) \). This policy is deterministic and fixed a
priori via depth-first search on the game’s forward model. To increase
human-likeness, the agent only jumps when necessary. This is realised
by always evaluating the idle action first in search. Also to further
human-likeness, the agent is subjected to a reaction time constraint, which
prohibits them to jump from the same position in which they landed.

- The lookahead hyperparameter \( n \). We have set \( n = 1 \) as this already
requires the anticipation of action consequences two steps ahead (cf.
Eq. 7.2). Given the fast pacing of RoboRunner, we would expect higher
values to exceed the anticipation capabilities of our human players.

- The agent’s sensorimotor dynamics model \( q(\hat{s}_{t+1}|\hat{a}_t, \hat{s}_t) \). We have set
this equal to the game’s forward model, as we at present do not model
any uncertainty on the side of the player.

We illustrate the three-step EBPXP procedure for a RoboRunner level in Fig. 7.6.
We did not calculate the reward for states that are either inaccessible nor
protected, indicated by the crossed-through squares in Fig. 7.6. While we
illustrate state-expected empowerment for all remaining states, the mean
state-expected empowerment as experience prediction is calculated only on
the states visited by our game-playing agent. A level’s loss is given by the
absolute distance between this PX prediction and a target mean:

\[
l := \begin{cases} 
|\mathcal{E}_{\text{target}} - \text{EBPXP}(g, p(a_t|s_t), n, q(\hat{s}_{t+1}|\hat{a}_t, \hat{s}_t), T)| & \text{if } g \text{ playable,} \\
\infty & \text{otherwise.} 
\end{cases}
\]  

(7.3)

Optimisation is performed for 300 generations or until \( l < 0.001 \). Our selection
is elitist in that the 5 best individuals are moved to the next generation
without modification. In addition, the 20 best individuals are subjected to
bitwise mutation with a rate of 0.05 and one-point crossover based on a uniform
distribution over the genotype length, and then added to the next generation.

We have generated and selected our conditions in three stages. We firstly (i)
instructed the genetic algorithm to produce level candidates over the whole
spectrum of prediction values identified through EBPXP. State-expected
empowerment for a 1-step lookahead can be at most 1, and is larger or equal to 0
(cf. Sec. 3.3). We evolved 21 levels with \( \mathcal{E}_{\text{target}} = 0, 0.05, 0.1, \ldots, 0.95, 1.0 \), real-
isising mean state-expected empowerment values \( \mathbb{E}[\mathcal{E}] \) in the range \([0.34, 0.92]\). These upper and lower optimisation bounds are not unexpected. We cannot
achieve zero empowerment as this would violate playability. For our AI agent
to reach the end of a level, platforms must be within jump distance, i.e. at
most four modules away from each other. Moreover, they must be at least
two modules wide for the agent to be able to land and jump off under our re-
action time constraint. We did not achieve the maximum mean state-expected
empowerment of one, as random mutation and crossover did not permit to
close all gaps within 300 generations. We secondly (ii) explored the generated
candidate levels to identify sensible conditions for our study. We observed
that levels with a mean state-expected empowerment below 0.65 had very short platforms and large gaps, and were thus almost unplayable. Levels with a mean above 0.95 in contrast had almost no gaps, and did not represent typical examples of an infinite runner. We consequently picked three values from the range of playable and typical levels: 0.65 for our low condition, 0.85 for high, and 0.75 for our tutorial level. We finally (iii) generated two instances of each condition and the tutorial level to balance for local structural differences that may affect PX. We ended up with eight unique level combinations. Fig. 7.7 shows the left third of both instances of each level condition with the AI play trajectory. Darker hues indicate lower state-expected empowerment for a player in that position. Protected and inaccessible states are crossed through.

To gain an intuition of empowerment in our testbed, consider the second low condition at the bottom of Fig. 7.7a. At position 58, the player could either jump to position 63, or keep running to position 64. Position 64 however is an abyss and empowerment is hence zero. In position 63 in contrast, both jumping and running would yield different futures and 1-step empowerment is thus maximum. The state-expected empowerment back at position 63 is consequently lower, because only jumping would allow the agent to get on solid ground and have future perceivable control.

7.5.4 Procedure

The experiment was run individually and under lab conditions. We first asked the participant to read a consent form and raise any potential questions. They then filled in a demographics questionnaire with the following items:

- Please specify your gender
  (Female, Male, None of the above, Prefer not to say)

- Please specify your age range
  (Under 12, 12-17, 18-24, 24-34, ..., 65-74, 75 or older, Prefer not to say)
We adopt Boyatzis’ (1998, p. 63) interpretation of a code as the most basic,

...
semantic feature of the raw data that can be assessed meaningfully with respect to the investigated phenomenon. We have specifically coded around the question of how players experienced a level, especially in relationship to previously played levels. A code dictionary was used between the experimenters to ensure consistency. Participants’ responses to targeted questions were analysed similarly to the think-aloud. Since this data is more reflective by nature, it allowed us to better understand some of the more immediate responses in the think-alouds and hence inform their coding. In the next step, we collated codes that occurred across participants into potential themes, and associated them with quotes from the interviews. In an iterative process, we reviewed and refined the codes and themes. We constructed an interpretation of each theme, accounting for the contribution of the associated codes. We eventually chose representative quotes from the transcript that evocatively illustrate our findings in the following report.

7.5.5 Participants

We recruited eight participants (5 male, 3 female) by opportunity sampling and incentivised them with chocolates. As justified earlier, this is a sufficient number of participants for an experiential vignette as exploratory, qualitative study. Six participants were aged between 25-34, and two belonged to age groups 18-24 and 35-44. Most were students in our local MSc and PhD games programmes. They were all English native speakers and avid players.

On average, our participants played videogames for 16.25 hours per month (SD = 5.29). Being asked about their favourite game genres, most named role-playing (6), followed by action (4), sports (2), adventure (2) and strategy (2) games. Individual mentions comprise horror, narrative, platformer, collecting and local co-op games. Our participants consider games enjoyable if they provide (optimal) challenge (6), have an immersive plot (5) and interesting mechanics (3), express a good idea (3), provide room for skill improvement (2), have clear graphics (1), interesting characters (1) and a good overall ‘feel’ (1).

From this data, we infer that our participants are skilled and have a diverse videogame playing experience. This is critical, as they should be able to get used to our testbed quickly and make some progress in the game in order for us to gather rich and comparable responses in the think-aloud.

7.5.6 Results

Our participants have reported different experiences in response to levels with different state-expected empowerment. The identified differences are notably not in the type of experience, which remained consistent between conditions, but in their strength and valency. We have identified the five major themes challenge, involvement, attention and engagement, and emotions as PX candidates that empowerment might allow to predict.

In the following report, we have paid particular attention to supporting our findings with quotes on both conditions that the players have been exposed to. If not referenced directly within a quote, we indicate the condition explicitly with a suffix. As annotations, we reserve (...) for omissions, and ... to indicate
7.5 EXPLORATORY STUDY

a pause in the utterance. Moreover, we put editorial additions and contextual comments into square brackets, with context information put into italics.

**THEME: CHALLENGE**

The most dominant theme in our interviews is *challenge*. Our participants frequently used words such as ‘difficult’, ‘tricky’ and ‘tough’ vs. ‘easy’ to express how challenging the current level is, and which we use as codes for this theme. They consistently considered *low* more *challenging* than *high*:

Boing, boing, ouh god it’s tricky! I see, yes! God is that...! [low] \(^{(1)}\)

This one [high] feels considerably easier. The frequency of gaps, and the size of gaps doesn’t feel as challenging [as for low]. \(^{(2)}\)

Being asked which condition they enjoyed more, one participant responded:

The second one [high] was quite easy. When it comes to the second one, it was a bit too easy I feel, ..., I got it in ... one life, I think? \(^{(3)}\)

Fig. 7.7 shows our different conditions next to each other, and hence highlights the impact of level structure on mean state-expected empowerment. We find that *low* has smaller platforms and more gaps than *high*. Our participants reported different challenges with respect to these structural differences, which we consider in the *sub-themes* of *physical* and *cognitive challenge*.

**SUB-THEME: PHYSICAL CHALLENGE**

Our participants articulated a *physical challenge* in their struggle to jump from one platform to the next in time, especially in quick succession. In our testbed, this and the other sub-theme of cognitive challenge can rarely be separated; one exception is given when the player hits a short platform, e.g. at position 57 in the lower level in Fig. 7.7a. Here, they must jump off again quickly to survive, and the imminent gap makes further reasoning unnecessary.

We have coded *physical challenge* via *reaction time* induced by threats, which has been articulated more frequently with respect to the *low* condition:

When the little, the small platforms are those that are challenging, because you have to bounce off them really quickly. [low] \(^{(4)}\)

Yah, I’m just going over the fact that I succeeded in one, when another one comes up, ... so it’s just about getting the time right. Maybe it’s to do with the rhythm. [low] \(^{(5)}\)

Whilst playing *high*, one participant felt lured into a physically challenging situation on a small platform although the level provides easier alternatives:

This is tricky, because you feel like you should jump in the middle, and I suspect when I jump on the middle one, I won’t jump off in time. [high] \(^{(6)}\)
### Sub-theme: Cognitive Challenge

Players experienced *cognitive challenge* on longer platforms where they face less instant danger but could jump off at different positions. We have coded *cognitive challenge* in terms of an increased demand towards *spatial reasoning* and *planning*, the potential to *devise strategies* and higher *decision pressure*.

Deciding on the best position involves *spatial reasoning* about jump and platform distances, which was almost exclusively mentioned for *low*:

> My spatial reasoning, to determine whether or not I can actually make that [jump]. So, it feels like ... it is asking me is whether my judgement of the distance, ..., whether I should jump sooner or later. It makes me question my ability to do that. [*low*]

If you chose the wrong thing, you would be left in a situation where you had to jump and land in a hole. So you wouldn’t jump at the earliest opportunity. [*low*]

One participant describes how they use their *spatial reasoning* while playing *high* to avoid a small platform in-between two larger and safer ones:

> Just cause they [the small platforms] are there, I like the safe bet. It feels quite safe to do that [jump on the small platform]. But, I can see that I can make it on the first jump. [*high*]

Participants leveraged their spatial reasoning to *plan* their actions, sometimes several steps ahead. This was mentioned almost exclusively for *low*:

> What I’m trying to do is to get myself into the situation where I can get my peripheral vision to see what’s coming up. (...) So I’m not just dealing what comes right at my doorstep. [*low*]

In response to whether they felt differently while playing the various conditions, players reported:

> There were certain bits I’ve come to, where there was the same distance of gaps but the platforms themselves were smaller, and there was a few ... they staggered, and I find it difficult to judge where to try and land to do another successful jump. [*low*]

> There were times were (...) you could choose to jump over either the first hole (...) or you could jump over more than one hole. And if you chose the wrong thing, then you would then be left in a situation where you had to jump and you had to land in a hole. So you wouldn’t jump at the earliest opportunity. [*high*]

Participants devised *strategies* to address the game’s cognitive challenges:

> So my new strategy for the small platforms is to just try to jump over them and not to land ..., well, if it’s possible to. [*low*]

Comparing their experiences of both conditions, one player responded:
As in ... I feel like [in low] I had to be more strategic, in terms of ... your jumping, when you jump, at what point you take off. And the other one [high] was a bit chilled. (14)

The need for the player to plan ahead, reason about distances and devise strategies while racing towards the next gap induces decision pressure. This was highlighted almost exclusively for low:

Doesn’t feel like I have a very big window to make the decision of whether to jump early or later. [low] (15)

The second one [low], it felt like they were a lot smaller, and less frequent, so there [in high] wasn’t so much of a challenge of ‘I have to make this decision now’.

**THEME: INVOLVEMENT**
Participants described an either active or passive gameplay experience. We use these two terms as codes for the player’s involvement in the game. We say involvement to capture an increased player activity as both, a response to the game’s demands and a voluntary act in the absence of such demands.

Our participants considered themselves passive exclusively when playing high. Being asked to compare the conditions, one responded:

(... the second one [high] seemed to have less of those really short platforms, and they had longer stretches of where you just didn’t, didn’t jump. So I kind of waited. (17)

Deliberating about their preferred condition in the targeted interview, another player emphasised their passivity when playing high:

Cause this one [high] has like too many ... where you’re just not jumping. It’s got quite a looong platform. So it’s kind of, just less interesting, than hard games. (18)

Participants generally considered themselves more active in low. Comparing the conditions, one player responded:

It felt like the second one [low] was the most difficult, as it asked more questions of me to try and make it to the end. So that was more involving. (19)

These ‘questions’ likely relate to the various factors of cognitive challenges elaborated earlier. A lack of threats in high did not discourage all participants from being active, with one jumping wildly on a long stretch without gaps:

Oh, I’m trying to do a double for fun. And another! [high] (20)
**Theme: Attention**

The degree of involvement in a game demands different levels of *attention*, a theme which we have formed from the codes of *concentration* and *focus*. Our participants referred to their *concentration* in an unspecific way, but expressed their *focus* with respect to particular game elements.

More than half of the participants mentioned the need to *concentrate* while playing *low*, expressed explicitly and through interrupted speech:

> Yeah, I’m having to concentrate a lot harder ... than before. And I’m, and I’m, ..., och, I’m keeping doing that. I’m keeping falling down this one. [low] (21)

Reflecting on whether the conditions felt differently, one participant noted:

> I felt like I had to concentrate harder on the second one [low], whereas the first one there were bigger spaces in-between each jump I had to make, so it felt comparatively more passive. (22)

In contrast to attention, our participants articulated their *focus* with respect to different game elements. Being asked whether they had spotted any differences between the levels, one player answered:

> I don’t know if the gaps were set on random or not. I was so focussing on jumping, I wasn’t actually looking at the sizes of them. I think maybe, ..., the spaces changed in the game as well. (23)

This participant did not refer to a specific condition. Another participant, in contrast, being asked whether the levels felt differently, responded:

> It’s a tougher game that you get into flow quite easily, because you’re quite focused on just jumping over the obstacles. [low] (24)

**Theme: Engagement**

When highlighting their attention, participants also expressed to be *engaged* with the game. We have coded this theme based on *determination*, the willingness to *re-engage*, and the ability to ‘zone in’.

More than half of our participants expressed their *determination* as the desire to succeed in a given level, or as disappointment when running out of chances. They mostly expressed this for *low*, but rarely also for *high*:

> Uuups! Damnit! I’m feeling a bit competitive now. [low] (25)

> Nooo! I’ve got no more lives? [low] (26)

> I’m gonna make these lives count. Now it’s personal! [high] (27)

Half of the participants mentioned the wish to *re-engage*, i.e. to play the same level again, for both conditions:

> (...) even though it is frustrating to constantly die playing a level I feel like I would’ve gotten there eventually. So it’s like you just play a little bit more. [low] (28)
Because level one [low] is so difficult, I don’t think I’d spend too much time on that if I kept on losing. (...) Now we’re talking about it, now I wanna go back and get it done [laughs].

Ohh rude! [disappointed] Can I restart? [high]

Our participants expressed to ‘zone in’, ‘be on a roll’, and be in ‘flow’. These accounts of engagement are too brief to warrant further differentiation, and we hence summarise them under the code zoning in. Being asked which condition they preferred, one player responded:

I preferred playing the second one [high]. I was on a roll. I was doing well, better than I was doing in this one [low].

Reflecting on how they felt when playing both conditions, another said:

[Low is] a tougher game that you get into flow quite easily, because you’re quite focused on just jumping over the obstacles.

**THEME: EMOTIONS**

Our previous impressions, e.g. on player’s perceived challenge, attention and engagement are reinforced by a large spectrum of emotions exhibited during the interviews. Almost all participants expressed pleasure when succeeding in the game and disappointment when failing, but some were also angry, tense and even anxious. These specific emotions are our codes for this overall theme.

Almost all participants showed pleasure when progressing in the game, but only in one instance while playing high:

Alright, I like the width of that one. I like the width of these ones. Yeah, nice and narrow. (...) Nice and narrow, it’s getting wider, uuuuh! ... double one, narrow one, ... oh, it’s much more satisfying not failing. [high]

Being asked which level they enjoyed more, the same player responded:

The one that I won [high], obviously! I guess you just get that payoff of pleasure, when you complete something.

Pleasure for low was only expressed non-verbally during play. Being asked which level was more enjoyable, another participant said:

So I think in a way I sort of enjoyed the challenges of the first one [low] even though I didn’t manage to complete it. The second one [high] felt a little too easy.

The majority of participants also experienced disappointment, usually about their performance (cf. quote 21), or about running out of lives to master a level. Disappointment was exclusively expressed while playing low:

Oh, fff... No more lives left. [low]

Uh! Oh nooo! Went tumbling... [low]
It’s okay, it’s okay, right: we’ve got this thick one [tense]! Aah (fell into hole)! See, there’s not enough space... [low] (38)

Most participants also experienced brief anger in both conditions, when loosing a live and especially when failing at the same position repeatedly:

So, cause I know what’s coming now, I played it enough times,..., I kind of, ..., bloody hell! [low] (39)

Ohh ... rude! Can I restart? [high] (40)

Almost all participants said or appeared to feel tense, especially when facing the unexpected and almost exclusively in low, as in quotes 39 and 38. Only once did a participant make their tension explicit in high:

It feels really tense, ... like I’m expecting something to happen. (41)

Our participants expressed relief after releasing tension. The most impressive example comes from a participant playing the tutorial level:

Sh! Uh oh oh oh, oh! Oh no! Ah! [Deep breath, exhaling] Puh! Release now, I’m really upset! (42)

In few cases and only for low, participants even expressed some (joyful) anxiety when anticipating and assessing upcoming situations in the game:

It’s this little one that gives me the heebie-jeebies. [low] (43)

Ah, oh no. I’m so scared []laughing]! [low] (44)

Assessing how the two conditions felt while playing, one participant said:

It [low] didn’t look different, I don’t think ... but it felt ... I was a lot more ... kind of anxious of this level. Ahm, it was kind a bit like adrenaline, because I knew ... well! Because I immediately lost a live ... (45)

MISCELLANEOUS THEMES
We have excluded three identified themes from this description. Participants reported remembering their actions and resulting performance, and recalling it again later. We summarised these codes into the theme of learning, but omitted it as it has been reported with similar strength for both conditions and hence does not help in discriminating the effect of empowerment. We also identified the theme of game design, based on players mentions of the game’s mechanics, aesthetics, complexity and story. They moreover related to the level layout as well as the shape of platforms and gaps, which we have summarised in a level structure theme. We have dismissed these themes as they do not represent an experience, and thus do not add to our research objective.

As is common in thematic analysis, the themes identified in this study are intended to stand alone as broad qualitative descriptions of the data. The method is not designed to look more causally at links between the themes. We hypothesise such potential links in the following discussion, but postpone a detailed examination to future work (Ch. 8).
We have developed our generic proposal to predict PX via IR calculated on gameplay simulations, and instantiated it in EBPXP, to facilitate more general and autonomous experiential control in PCG as a core area of videogame AI. Although our model is not limited to this application, we have evaluated it qualitatively as a component of a procedural content generator. The thematic analysis of player think-alouds supports our prediction PD.6 that game levels which are predicted to yield different PXs by EBPXP indeed evoke different experiences in human players. We moreover identified a range of candidate PXs which EBPXP could potentially predict; for our model to be of practical value, these must be further investigated and confirmed via a future quantitative study. In the following discussion, we relate the themes identified through our thematic analysis back to PX research to form a working hypothesis for this future quantitative study. We then highlight the limitations of our approach, and discuss its potential use in CC.

Perceived challenge represents the strongest theme in our study; players have consistently reported levels of the low condition to be more challenging than high, exposing it as a candidate PX which empowerment may allow to predict. We can distinguish different types of challenge, depending on which player abilities are being addressed (Denisova, Guckelsberger & Zendle, 2017). In our study, players experienced physical challenge which is understood to address a player’s physical limitations to interact with a game, e.g. the speed and accuracy with which they can perform required actions (Cox et al., 2012). Moreover, our participants have experienced cognitive challenge which is thought to address the player’s cognitive capacities, and the speed and accuracy of their prediction and problem solving facilities (ibid.). Players have not related to any emotional challenges that typically result from emotionally salient material, the use of strong characters and a captivating story (Cole, Cairns & Gillies, 2015; Bopp, Opwis & Mekler, 2018). This is not surprising, given RoboRunner’s genre and relatively simple story. In the development of a challenge questionnaire after and independently of this study (Denisova et al., 2020), we have identified the additional type of decision-making challenge. It arises when a player has to make choices that are difficult or lead to a regrettable outcome. Retrospectively, we acknowledge that this could be separated from the theme of cognitive challenge (e.g. Quote 8).

A theme which has previously received little attention by the games user research community is involvement; we have associated it with our participants’ active or passive play experience in the low and high condition, respectively. In the absence of any dedicated research on this experience, we relate to two game genres which demonstrate it in extreme form. Firstly, Juul (2010) refers to a reduced demand for player activity as characteristic of casual games. Moreover, idle games (Alharthi et al., 2018), also called incremental and clicker games, typically feature a continuous decrease of the player’s involvement: being initially in strong demand, a player must gradually replace their own activity with more efficient substitutes to progress in the game, eventually reducing their involvement to a minimum or making them obsolete altogether.

Our analysis moreover suggests that the low condition demanded more player attention than high. We have coded this theme in terms of concentration
and focus; Cutting (2018) in contrast, based on prior research by Lavie et al. (2004) and others, understands players to pay more attention to a game if they are less easily distracted. In our study, participants playing the low condition were less committed to thinking aloud, which could be interpreted as a separate, distracting task. Our conceptualisation of attention thus aligns with his operationalisation. Moreover, Cutting identifies ambiguity in existing research on the relationship of player attention and their cognitive load: more specifically, his experiments do not support Lavie et al.’s (2004) finding that increased cognitive load reduces attention. We even observe the opposite: reports of cognitive challenge suggest that the low condition comes with a higher cognitive load, but it yet seems to attract more player attention.

We finally contextualise engagement as the fourth identified theme. We considered players engaged when they appeared determined, willing to reengage and reported to ‘zone in’, ‘be on a roll’, or be in ‘flow’. In contrast to the other themes, no specific condition stood out as particularly engaging. This is not surprising, as engagement is one of the most complex PXs (Boyle et al., 2012). Based on a review within and beyond videogames, O’Brien and Toms (2008) define engagement as ‘a quality of user experience characterized by attributes of challenge, positive affect, endurability, aesthetic and sensory appeal, attention, feedback, variety/novelty, interactivity, and perceived user control’ (ibid., p. 941). In the development of an engagement questionnaire, Brockmyer et al. (2009) consider the underlying concepts of immersion, presence, flow, psychological absorption, and dissociation. Our coding hence only covers very few high-level attributes of this PX. At the same time, engagement rests on other experiences identified in our study, such as challenge and attention.

We deem engagement too complex and high-level to be well predicted by empowerment as IR. This is also supported by its inconsistent association with our conditions in the think-alouds. For similar reasons, we discard affect, the last identified theme, as candidate experience to be predicted through EBPXP. We moreover believe that the relationship between the state-expected empowerment reward used in our model and the remaining three identified experiences of challenge, involvement and attention is only intermediate, as these experiences are still rather complex. Our study yet helps us to identify more low-level experiences that empowerment could predict with better accuracy: informed by our discussion of related game design concepts and games user research findings in Sec. 7.4, we have sought experiences that commonly influence all candidates identified in the think-aloud. We arrive at the following hierarchical working hypothesis:

State-expected empowerment allows to predict the foundational experiences of effectance, outcome uncertainty, and perceived control.
These have mediating effects on a player’s perceived challenge, attention and involvement, which in turn influence high-level goal experiences such as engagement, enjoyment and affect.

We adopt the notion of goal experiences (Cairns, 2016) for the high-level experiences that players are explicitly seeking in gameplay, and foundational experiences (Power et al., 2019) for the low-level building blocks that these are formed of. We are not surprised that players did not report these foundational
experiences explicitly, as this is not what they are primarily looking for. We describe future work on evaluating this hypothesis in Ch. 8.

Our study only marks the beginning to investigating intrinsic reward-based PX prediction, and has several limitations. Firstly, think-alouds on different level conditions can reveal the breadth of players’ experiences, and coarsely which condition they were caused by; however, they cannot reveal their intensity, and hence which PX our IR is best suited to predict. Moreover, the small number of participants, while being sufficient for an experiential vignette study, and also their limited diversity, e.g. in terms of playing experience and cultural background, prohibits any definite conclusions. These limitations arise naturally from this study type and were anticipated; our aim was not to make definite statements, but to use the rich, qualitative data provided by our participants to explore and constrain the choices for follow-up studies. We map the path ahead in our future work Ch. 8.

Secondly, many commercial games are more complex than RoboRunner. As a one-button, two-branching factor game, the space of possible, successful gameplay manoeuvres is small. There is little opportunity for self-expression and personalisation, and for different player types to evoke fundamentally different experiences. The game for instance offers little to to spur and satisfy the curiosity of an explorer (Hamari & Tuunanen, 2014). This leaves open whether EBPXP and our more generic approach would perform in a similar way as in our study; it is for instance unclear how the accuracy of predictions would be affected by games that allow the player to progress non-linearly, or that elicit many overlapping experiences. We decided to develop RoboRunner instead of using a commercial game to warrant experimental control throughout, from content generation to gameplay evaluation. Our simplistic design allowed us to disregard different play styles as variable in our experiments, and to put more emphasis on experience prediction than on the construction of a complex player model. Moreover, it enables us to polish the testbed sufficiently so that participants did not question its authenticity, and did not become distracted by potential flaws in its design or implementation. This simplicity also allowed us to explain the reported experiences based on the interaction of few game elements. Crucially, RoboRunner is not a mere abstraction of actual games, but many commercial games of a similar kind and complexity exist, especially for mobile devices. Within this niche, our results are of immediate relevance.

Thirdly and finally, our study does not tell us how well our generic approach or specific model would generalise to other games that are not necessarily more complex, but may e.g. represent a different genre. At present, mainly the formal properties of IR, as argued in Sec. 7.3, and the related work surveyed in Sec. 5.2.1, support that such generalisation is possible.

These points concern what can and cannot be inferred from our present study. We next consider limitations to the underlying approach in its present formulation. We distinguish implications of the choice of intrinsic reward (IR) for the PX model, and the AI agent for its assessment.

A specific IR can only yield useful predictions in games that afford fluctuations of this reward. Different states in RoboRunner for instance give the player more or fewer future options, which are captured in a diverse empowerment landscape. In contrast, we can imagine a narrative game in which
the player always has the same number of choices, leading to an equal amount of non-overlapping, distinct changes to the narrative. No matter what trajectory a player would take through this game, EBPXP would always predict the same experience, as the underlying empowerment landscape is flat. Yet, the game likely conveys many other experiences that are not related to e.g. their perceived effectance or control. We suspect that these can be captured by other IRs (Sec. 2.2.4). While employing IR in PX prediction likely increases generality, the latter is yet constrained by the nature of the chosen reward.

A separate threat to the accuracy and generality of our approach is what made it more independent from people and thus more flexible in the first place: the AI agent used to simulate human gameplay. The accuracy of our prediction is strongly influenced by how closely the behaviour of this AI agent resembles common human play. Most existing work on general game-playing relies on human designers to specify a game’s goals as target for play. This reliance on people and game-specific goals reduces the generality of these models. In addition, they often yield behaviour that is not human-like. On the opposite, human-like AI agents often require human player data for training, which also impedes on their generality. We address the limitations of both our study and our approach as part of future work (Ch. 8).

While our study is specific to games, it can inform the application of intrinsic reward-driven experience modelling to other interactive artefacts. As motivated in Sec. 7.1, this could advance CC evaluation by overcoming several shortcomings in present models. We expect EBPXP to predict similar foundational experiences as identified through our study for the interaction of an audience with e.g. a piece of installation art, or an interactive film. For instance, it could provide an estimate of the effectance which a kinetic sculpture would afford to a generic audience. These insights could be used, potentially with the help of domain-specific theories, to predict a range of goal experiences: how much the sculpture would challenge a person’s cognitive or performative abilities, how engaging it would be, and how well it would hold their attention. These predictions could then contribute to driving an artificial system to autonomously create and evaluate new sculptures that are valued by a human audience. Realising these new applications requires us to tackle some hard questions in future work, as discussed in Ch. 8.

In this chapter, we have proposed and evaluated an end-to-end experience-driven PCG approach. Following Yannakakis and Togelius’ (2011) taxonomy, it combines global search-based content generation on a direct content representation with content evaluation based on a gameplay- and model-based PX model and simulation-based content quality assessment. Our core innovation in Sec. 7.3 and 7.4 concerns the evaluation component, and we hence primarily contribute to player modelling as a separate area of game AI (Yannakakis & Togelius, 2018, pp. 259–260). As such, it can benefit PCG as motivated in Sec. 7.2, but also other applications within and beyond game AI. Our exploratory study in Sec. 7.5, subject to the limitations pointed out earlier, supports that IR can be used predict people’s experience of games as interactive artefacts, thus cautiously affirming our research question RQ.9. In the next chapter, we consider future work across all contribution of this thesis.
Part IV

OUTLOOK
FUTURE WORK

In this thesis, we have made four central contributions to help shaping the future of CC and videogame AI by means of computational intrinsic reward (IR) and models of intrinsic motivation (IM). Here, we present a selection of the many promising directions that our work could be taken in the near future and in the long term. To avoid redundancy, we describe next steps on expanding each of our systematic reviews jointly in Sec. 8.1. We then propose future work on advancing our applied contributions in Sec. 8.2 and 8.3 with respect to the conducted studies, the introduced model, and its application to advance CC beyond game AI. Sec. 8.4 concludes this chapter with examples on how our applied work could be consolidated.

8.1 Systematic Reviews

In Ch. 4 and 5, we conducted two systematic reviews to answer why researchers have embraced IR and IM models in CC and videogame AI (RQ.3 and RQ.5) and how such rewards and models have been applied in both domains so far (RQ.4 and RQ.6). To answer these questions with respect to CC, we have traced relevant work from a set of traditional publication venues, and filtered it based on our working definition of both CC (Sec. 4.2.1) and models of IM (Sec. 2.2.3). We ended up with 29 related work items from as early as 1998 to 2018. We have applied a similar systematic approach to tracing related game AI work. We filtered the identified candidates based on our working definition of IM models, complemented by Yannakakis and Togelius’ definition of videogame AI (2018, p. 4) and Juul’s (2003) definition of videogames (Sec. 5.1). Our selection is biased towards work that uses IM for the benefit of games, putting less weight on contributions that use games as a benchmark for artificial general intelligence. Our final selection comprises 11 related work items dating from 2006 to 2019. We have identified the same reasons to embrace IR and IM models in both bodies of related work, and distilled them into a typology as answer to RQ.3 and RQ.5. Moreover, we have identified 12 abstract applications of IR and IM in CC, linked to (computational) creativity theories and informed by creativity research findings on the relationship of IM and people’s creativity (Sec. 4.1). We have also identified 11 abstract applications of IR and IM models in videogame AI. These are linked to the four core domains of game AI (Yannakakis & Togelius, 2018, pp. 262-264), and informed by game design and games user research findings on what makes games intrinsically motivating for people (Sec. 5.1). We have arranged these applications in two separate typologies as answers to RQ.5 and RQ.6. In Sec. 5.2.1.3, we identify a strong conceptual overlap between applications of IR and IM models in CC and videogame AI, thus answering RQ.7 and framing related game AI work as part of computational game creativity (Liapis, Yannakakis & Togelius, 2014). We discuss how to expand these reviews in future work, and point out underexplored areas in both domains.
Our systematic reviews are strictly speaking incomplete, as they do not include our own contributions to both domains. As a first step, we must thus integrate and contextualise our work on IM in CC and game AI directly into the respective review narratives and typologies. Moreover, we want to expand our CC review specifically by referring to related work in other fields which does not address creativity explicitly and was hence excluded based on our working definition of CC, but is yet relevant for advancing the field’s central goals. We aim to include more work from machine learning such as Mahadevan’s (2018) proposal of ‘imagination machines’ which shares many common themes with other CC publications. Moreover, we are keen to draw connections to theories and applications of IM to adaptive behaviour, artificial life, open-ended learning and developmental robotics; related contributions in these domains implicitly focus on creative behaviour, addressing questions of mini-c creativity in development and adaptation, and can thus complement CC’s research focus on big-c, artistic, artefact-oriented creativity. These connections have previously been emphasised by Schmidhuber (2006), Saunders (2012), as well as Aguilar and Pérez y Pérez (2015), amongst others. We consider this an important effort in pointing out cross-disciplinary connections and promoting CC as a joint research effort (cf. Sec. 4.2.1).

We are particularly committed to complementing our reviews with an account of underexplored but promising areas of future research on IM. Mapping these areas constitutes future work in itself, and we only name a few examples here that we consider particularly crucial to be followed up. Existing work can commonly be characterised by its strong focus on variations of curiosity as a specific model, and an ignorance with respect to IM as a more general family of motivational models with common properties. We thus consider it most important to introduce other IRs and models of IM to both CC and game AI, and to investigate how their common properties can be leveraged to tackle challenges in both domains. In Sec. 8.2 and 8.3, we make concrete proposals for embracing IR other than empowerment in the applied contributions of this thesis.

We also observe a common trend in CC and game AI towards open-endedness: Cook and Colton (2018b) propose overcoming the one-shot generation that dominates CC and to engineer systems that pursue larger goals, draw from a wider (temporal) context, and thus have more presence in the world. They illustrate this with the vision of turning the automated game designer ANGELINA into ‘always-on’ systems, and hence bridge to computational game creativity and game AI. Closely related, Gaina, Lucas and Pérez-Liébana (2019) propose a ‘forever gameplayer’ that could function as a component of such an open-ended game design system. Both visions are limited in their potential autonomy and presence by relying on extrinsic reward. We believe that they could be substantially advanced by the use of IR, as it affords an embodied and subjective presence, and can be used within a model of IM which gives rise to open-ended development (Sec. 2.2.3).

We next point out two underexplored research areas in CC and game AI individually. Our review in Ch. 5 reveals PCG as the core game AI domain (Yannakakis & Togelius, 2018, pp. 262-264) which has so far been addressed the least with models of IM. We propose to explore IR in RL-
8.2 intrinsically motivated social co-creativity

In Ch. 6, we motivated the need for more general, co-creative artificial agents capable of either supporting or challenging their interaction partner in human-computer co-creativity more generally and in NPC AI specifically. Models of IM have been successfully used to increase the generality of agents in both domains, but existing models cannot induce stable social dynamics. Drawing on this prior work, we have informally introduced social models of intrinsic motivation to overcome this shortcoming. With a focus on driving the behaviour of companion and adversary NPCs, we have introduced coupled empowerment maximisation (CEM) as such a social IM model informally and formally. In two exploratory, qualitative studies based on the novel method...
of observational vignettes, we have probed the capacity of CEM to drive the behaviour of general, believable NPCs that either support or challenge the player as companion and adversaries, respectively. Through the lens of NPC AI, our findings affirm this chapter’s research question, ‘Can we use a model of intrinsic motivation to engineer general and social co-creative agents?’ (RQ.8), within the studies’ limitations. Here, we propose future work to overcome these limitations and to investigate this question further (Sec. 8.2.1).

We then outline directions for extending and applying CEM in and beyond game AI (Sec. 8.2.2). We finally return to our motivation and consider how our work could be applied to other creative domains in the future and advance central goals of CC (Sec. 8.2.3).

8.2.1 Directions for Further Study

Our exploratory, qualitative studies in Sec. 6.5.1 and 6.5.2 have provided rich insights into the nature of CEM-driven NPC behaviour, helped us in identifying different types of co-creativity, and inspired future applications of our approach, amongst others. To apply this principle in actual games and to advance the study of creativity in the interaction of NPC and player, we must complement these qualitative insights with quantitative measurements. Most importantly, we must assess how strongly CEM-driven NPCs support and challenge the player, and quantify the diversity of the respective behaviour as a separate determinant of NPC believability (Sec. 6.2 and Tbl. 6.1). We propose to do this through objective and subjective quantitative measures, the latter applied to assessing the perspective of both players and designers.

The most relevant work for the subjective assessment of social interaction dynamics in games is Hudson and Cairns’ (2014) ‘Competitive and Cooperative Presence in Games’ questionnaire. However, their scales capture a player’s perception of opponents and team-mates, rather than that of adversaries and companions. These character types are assumed to work towards the same goals as the player (Sec. 6.2), which is not necessary for behaviour to be considered supportive or adversarial. Rather than employing this sub-optimal instrument, we suggest using it as inspiration for the future development of a new and well-validated questionnaire to measure players’ perception of companions and adversaries. We also propose to include a scale that assesses people’s perception of the novelty, diversity and surprisingness of character behaviour, as this is missing from existing work, and could contribute to the assessment of character believability. When devising such a questionnaire, we must consider that certain experiences, in particular affect, may be by nature relative and could thus be more reliably assessed through ordinal, rather than nominal and interval measurement methods (Yannakakis, Cowie & Busso, 2017, 2018). We would expect such an instrument to be of considerable use beyond this immediate context. It could for instance be used in game development to assess player’s perceptions of companions and adversaries, or in games user research to investigate how this perception is affected by certain manipulations of the player and the game.

Togelius et al. (2013) argue that ‘participatory observation, where the human assessing believability takes part in the game, is prone to distortion’,
e.g. due to learning effects and self-deception, the intrusiveness of self-reports during gameplay, and their sensitivity to subjective memory limitations if assessed post-play. We suggest to follow Togelius et al.’s proposal and complement a participatory, subjective assessment of NPC believability with a Turing test (Turing, 1950) like variant in which external observers rank the performance of human and non-human controlled agents without playing the game themselves and without knowing who is in control.

A candidate measure to objectively assess the strength of support and antagonism induced by CEM is the performance difference resulting from playing with and without a CEM-driven NPC. This performance could be assessed on game score or other performance indicators. At present, we are not aware of a similarly natural candidate to quantify the novelty, diversity and surprisingness of NPC behaviour. The arguably biggest challenge is to chunk potentially overlapping, variable-length sequences of atomic actions into meaningful behaviours. Existing work relies on manually annotating such action sequences in advance (Merrick & Maher, 2006), or training a classifier on observed typical behaviours (Soares & Bulitko, 2019). It might be worthwhile to explore intrinsic novelty rewards (e.g. Schmidhuber, 1991; and Sec. 2.2.4) to facilitate the chunking of action sequences and novelty detection without supervision. Moreover, predictive information (Bialek, Nemenman and Tishby, 2001; and Sec. 2.2.4) could be used to measure the complexity of a behavioural process.

Irrespective of the measurement type, the evaluation of NPC believability should be conducted from an embodied perspective: we must not only measure the impact of the AI controller on the NPC’s believability, but also how believability is affected by the game world and the way in which it is perceived and can be acted upon by the character. Following a brief note by Togelius et al. (2013), Camilleri, Yannakakis and Dingli (2016) complement the traditional controller- with a game content-centred perspective. They confirm through a user study that NPC believability is substantially shaped by the game world. We expect this to hold even more for intrinsically motivated NPCs, whose rewards are strongly influenced by their embodiment.

Our vision is to engineer NPCs that could realise believable companion or adversary behaviour in a wide range of different games and in response to different players, with little or no changes to the underlying motivational principle. As an early proof-of-concept, our two studies only provide limited evidence of such game and player generality (Togelius & Yannakakis, 2016).

To support CEM’s game generality more strongly, we must study it beyond our own modifications of a game in a wide range of existing games from different genres and with diverse mechanics. To support CEM’s player generality, these games should not be played by the potentially biased experimenters, but by a representative selection of players with varying skills.

We propose to investigate player generality in future work by controlling the player avatar with game-playing agents of varying sophistication (cf. Nielsen et al., 2015), or by recruiting a representative range of human players. Studying CEM’s game generality may be more difficult. Commercial games typically do not allow for the straight-forward integration of CEM and do not afford sufficient experimental control to evaluate its workings. We must hence resort to game AI benchmarks, but only few cover the multi-agent case. Moreover, these assess exclusively cooperative and competitive behaviour, rather than
antagonism and support. Finally, these benchmarks are considerably less rich than commercial games, and might offer very little in terms of an empowerment gradient. We still propose to appropriate these existing benchmarks to explore their potential for investigating CEM-driven NPCs. As a first next step, CEM could be evaluated on the competitive and collaborative games of the general video game AI (GVGAI) two-player track (Gaina, Pérez-Liébana & Lucas, 2016). These games are discrete and come with a forward model. The latter enables us to evaluate the simplified CEM formulation (Sec. 6.4.3.3) separately from the model acquisition problem. The Multi-Agent Reinforcement Learning in Malmö framework (Perez-Liebana et al., 2019) represents an interesting candidate to evaluate more general and scalable future versions of CEM (Sec. 6.4.3.2). The competitive and collaborative games run within Minecraft (Mojang & Microsoft Studios, 2009) and are hence continuous and in 3D. They do not come with a forward model, and CEM must thus either incorporate model acquisition or leverage a model-free method to calculate empowerment, as discussed below.

8.2.2 Improvements to and Potential Applications of Our Approach

We have motivated CEM based on its potential to drive robust, supportive or antagonistic NPC behaviour even if the game changes. A certain configuration of hyperparameters of the simplified CEM model, i.e. the empowerment weights $a = (a_C, a_P, a_{CP})$ and the lookahead $T$, already warrants a high degree of robustness and hence generality. This is supported by our studies, in which one default configuration yields sensible behaviour across all conditions. However, if one specific empowerment reward is weighed too strongly, it might impose a bottleneck in the character’s movement towards a game’s goal states, and result in them getting stuck in local coupled empowerment maxima. We believe that future improvements should aim at eliminating the need for manual hyperparameter tuning altogether to further enhance CEM’s generality and avoid characters acting towards local optima. We distinguish two complementing directions of research.

We propose to automatically tune the hyperparameters based on the simulated interaction of the NPC with a game-playing agent controlling the player avatar. To yield sensible NPCs, the cost function must quantify believability as discussed in the previous section, especially the strength of support and antagonism. Alternatively, actions in the CEM policy could be selected based on the satisfaction of an empowerment constraint hierarchy, rather than the linear combination in the action-value function (Eq. 6.45). Each such hierarchy would model a certain NPC persona. For instance, a self-sacrificing companion would always pick the action that maximises the player’s empowerment. If multiple actions fit this target, the NPC would then select an action subset that optimises the NPC-player transfer empowerment, followed by their own empowerment. We can also imagine an NPC who selects actions to first optimise their own empowerment, followed by the player’s and then the NPC-player transfer empowerment. This NPC would likely be perceived as a more cautious companion which only supports the player if this poses no
threat to themselves. This could complement the previous approach, in that only the lookahead remains to be tuned automatically.

A CEM-driven NPC may also get stuck in local optima due to their use of greedy action selection. We hence propose to determine action-value based on aggregated coupled empowerment, e.g. through action-value, policy gradient or actor-critic RL (Sutton & Barto, 2018). Crucially, selecting actions based on a future coupled empowerment reward over multiple timesteps is not equivalent to calculating the coupled empowerment reward for a longer lookahead $n$. More research must be conducted on the trade-off between temporally extended action-value and the hyperparameter $n$.

We deem improvements to the scalability of CEM critical to increase the behavioural complexity of the controlled agents and to deploy them in more complex domains, in particular commercial games. We believe there is much promise in informing the future development of an approximate, efficient version of CEM by recent approximations of single-agent EM in RL.

Two components contribute to CEM’s present exponential complexity (Sec. 6.6): the (i) calculation of the $n$-step predictive factor and the (ii) calculation of empowerment as mutual information maximisation. Rather than tackling (ii) with the exact but exponentially complex Blahut-Arimoto algorithm (Arimoto, 1972; Blahut, 1972), recent approximations all sample a variational bound (Blei, Kucukelbir & McAuliffe, 2017) on the mutual information, parametrised as a neural network. Existing work can be distinguished in the solution to (i): while model-free approaches sample the variational bound straight from agent experience, model-based approaches optimise the bound with a learned model, and use experience only to find the empowerment maximising policy. A model-free approach has been used by Mohamed and Rezende (2015) to approximate open-loop and by Gregor, Rezende and Wierstra (2017) and Binas, Ozair and Bengio (2019) to calculate closed-loop empowerment (Sec. 3.3), in both the discrete and continuous domain. Karl et al. (2017) in contrast use a separately acquired model to optimise continuous open-loop empowerment.

We propose to explore the approximation of CEM with a model-based approach, as model-based RL has greatly improved in recent years (Wang et al., 2019), comes with a higher sampling efficiency, and allows us to consider model acquisition separately from the coupled empowerment reward calculation. Another promising avenue for future research is to increase the lookahead in CEM by utilising macro-actions, e.g. in the form of options in hierarchical RL. Finally, more informed, low-entropy policy models, e.g. acquired through inference, could not only improve the quality of behaviour, but also reduce the branching factor in the calculation of the $n$-step predictive factor.

Our long-term vision is to gradually relax the assumptions of the simplified CEM model in Sec. 6.4.3.3 and arrive at the generic formulation in Sec. 6.4.3.2. This requires tackling several state-of-the-art machine learning challenges, but it should be possible to treat them separately. Each relaxed assumption unlocks new features of CEM-driven NPCs to benefit videogames. For instance, not every game provides access to a forward model, or the model might be very complex and hence expensive to sample from. If we manage to relax the assumption of fixed parameters and perfect sensorimotor models, the NPC can learn only the proportion of the actual forward model that is relevant for the calculation of the different empowerment variants.
As a second example, learning the model of the player sensory dynamics, environment dynamics and policy would improve coordination and hence the quality of support or challenge. It would also enable an open-ended interaction with players, in which the CEM-driven NPC adapts their behaviour to changes in e.g. the player’s strategies, thus expanding their behavioural diversity. Such behaviour is desirable: Emmerich, Ring and Masuch (2018) for instance find through an online survey that players deem it ‘exciting if the relationship between [their] character and the NPC evolves during the game’ (ibid., p. 149) \((M=3.53, SD=0.80)\). Similarly, Yannakakis and Hallam (2005) argue that videogames become more interesting when the player can engage with NPCs, e.g. opponents, that adapt on-line. Progress on this requires the use of powerful inference mechanisms and theory of mind models (e.g. Rabinowitz et al., 2018; Raileanu et al., 2018). As a last example, not assuming the latent environment state to be known but inferring it instead could increase the believability of CEM-driven NPCs in incomplete information games (e.g. Bard et al., 2020), where the inference of hidden information is a core gameplay element. More generally, relaxing the model assumptions is crucial for applications of CEM outside videogames as tightly controlled environments, e.g. in human-robot interaction (Salge & Polani, 2017).

We put forward two applications of CEM and social IM models (Sec. 6.3) more generally which we would particularly like to see explored. Firstly, we encourage the application of such models to automated game design. Existing systems such as ANGELINA (e.g. Cook, Colton & Gow, 2016a, 2016b) already use NPCs that realise a few generic, pre-made behaviours. While these NPCs work in a large range of produced games and thus do not limit a system’s expressive range (Smith & Whitehead, 2010), they may be perceived as inapt and boring. In principle, automated game design systems can generate any kind of NPC behaviour through code synthesis; the challenge is to generate a custom-tailored NPC that shows sensible behaviour that is specific to a given game. Social models of IM such as CEM represent one means to alleviate this challenge, in that they leverage all functional interactions that the present game design affords for sensible behaviour. We hope to see e.g. engaging and believable, CEM-driven adversaries in automatically designed games soon.

Our second proposal addresses the games industry requirement for NPC AI to be predictable (Yannakakis and Togelius, 2018, p. 14; and Sec. 6.6). To mitigate this requirement and leverage more of CEM’s benefits, we suggest employing the motivational model during development. More specifically, CEM could be applied in game prototypes to inspire designers with new and surprising supportive or antagonistic behaviours to drive the engineering of traditional, predictable NPC AI. These NPCs could either be hand-authored from observations of CEM-driven behaviour, or they could be learned from these observations in an unsupervised fashion, e.g. in the form of decision trees, or the utility vectors and policies of procedural personas (Holmgård et al., 2014a, 2014b), which can be inspected and adjusted prior to deployment.
8.2.3 Next Steps in Advancing Computational Creativity

For the development of CEM, we have drawn on the strengths and weaknesses of existing approaches to IM-driven human-computer co-creativity and NPC AI. As argued in Sec. 6.2, we understand this as a contribution to computational game creativity (Liapis, Yannakakis & Togelius, 2014). Here, we advocate two directions for future work on advancing the goals of CC more generally through social models of IM. The first direction is to leverage NPC AI as a laboratory to study computational (co-)creativity in behaviour. At present, creativity in behaviour is barely addressed in CC, and videogames present a unique framework for inquiry. A game’s goals are typically clearly identifiable, and due to games being autotelic activities (Salen & Zimmerman, 2004, pp. 332-333), at least some of these goals are constrained to a game’s magic circle (cf. Huizinga, 1950, p. 10; and Sec. 5.1.1). Hence, games offer us unambiguous reference points to assess the value of creative behaviour. As entire worlds within our reality, games allow for the investigation of systems theories of creativity (e.g. Csikszentmihalyi, 1988) under tight experimental control, and for the assessment of novelty within a closed frame of reference. The focus on NPCs enables the consideration of creativity from many different angles (Sec. 4.2.1): the autonomous creativity of the NPC, the co-creativity between the NPC and the player as well as the designer; and the NPC’s creativity support towards human design and gameplay.

Our studies in Sec. 6.5.1 and 6.5.2 represent first steps in addressing open questions on the creative behaviour of artificial systems. We urge researchers to consider future work from different perspectives (Sec. 4.2.1). From a cognitive and systems theories perspective, it would be interesting to see how changes to the game world and character abilities impact the creativity of a CEM-driven adversary. Moreover, we would be intrigued to find out how a player’s exploratory and transformational creativity (Boden, 1990/2003; Wiggins, 2006a, 2006b) is influenced by the interaction with such an autonomous, supportive or antagonistic co-creative partner. From an engineering perspective, it would be beneficial to see which choice of IR in social IM models makes the NPCs more creatively autonomous, or increases a player’s and/or designer’s perception of their creativity. This framework crucially does not rely on a human interaction partner; we can equally study the interaction between different NPCs as in social creativity systems (Saunders & Bown, 2015).

A second direction for future work is to return to our original motivation in Sec. 6.1 and employ social models of IM in other CC domains to overcome generative impotence (Kantosalo & Toivonen, 2016), alleviate the generality-quality trade-off, and engineer autonomous CC systems capable of supporting and challenging their interaction partners in open-ended co-creativity. This requires researchers to explore which IRs as the basis of such social IM models yields meaningful behaviour in the respective domain, just as we did for videogames in the development of CEM (Sec. 6.4). There are many IRs to choose from, with some discussed in Sec. 2.2.4. Crucially, a specific reward is not necessarily restricted to one creative domain: empowerment as the foundation of CEM could in principle be employed in any domain where increasing and decreasing the partner’s options and influence is perceived as support and antagonism, respectively.
8.2 Intrinsically Motivated Social Co-Creativity

Figure 8.1: Thought-experiment on CEM-based antagonism in Curious Whispers (Saunders et al., 2010). The CEM-driven robot (blue) decreases the empowerment of a coupled (orange, dashed line) robot by playing a tune which interferes with their performance to others.

Figure 8.2: Thought-experiment on CEM-based support in Curious Whispers (Saunders et al., 2010). The CEM-driven robot (blue) arranges blocks to shield their coupled (orange, dashed line) partners’ rehearsal from interference by others, hence increasing their empowerment.

To jump-start this exploration process, it might be helpful to investigate social IM models in existing intrinsically motivated, co-creative or social CC systems, as surveyed in Sec. 4.2.2. These embodied systems were originally designed to provide rich input to an IR, and are typically situated in a complex world that affords and requires open-endedness. Moreover, they have already been evaluated for a specific kind of IR. We can put this candidate into a social IM model, and use the existing observations for a comparison with the previous implicit reward alignment approach (Sec. 6.1).

For instance, it would be fascinating to instantiate a model of coupled novelty maximisation in Curious Whispers (Saunders et al., 2010). Likewise, we would
like to see CEM applied to realise support and antagonism in this society of tune-generating robots. We briefly motivate this with thought-experiments. An antagonistic, CEM-driven robot could reduce the empowerment of other robots by playing a tuned which interferes with theirs, and hence makes it impossible for them to differentiate between their generated tunes. Consequently, a robot listening to their performance would not be able to pick up the original tune (Fig. 8.1). If the coupled robot was able to express different tones, they could recover their empowerment by switching to a spectrum which the antagonistic robot could not disturb. A supportive CEM-driven robot could modify the environment to improve the empowerment of their coupled peers: if there were movable blocks, they could shift them into a position which shields their coupled peer from interfering robots, and hence improves their rehearsal (Fig. 8.2). We anticipate that these social dynamics could give rise to interesting emerging effects that influence the differentiation of artefacts, formation of cliques, etc., thus supporting the goal of Curious Whispers to investigate systems theories of creativity. Guckelsberger et al. (2016) present further thought-experiments, and also consider equipping previously extrinsically motivated CC systems with IM models.

8.3 INTRINSIC REWARD-BASED EXPERIENCE PREDICTION

Our focus in Ch. 7 has been on modelling the human subjective experience of interactive artefacts. We have pointed out the importance of this form of evaluation for the (perception of) creativity, the autonomy of CC systems, and for unleashing the potential of PCG as our game AI application domain. We have argued that these goals are severely limited by existing techniques, which must involve people when the artefact or the generator changes, rendering them inflexible and constraining their generality. We have proposed to overcome these and other shortcomings by estimating the human experience of interactive artefacts via IR and AI agent simulations. Focussing on PCG, we have instantiated our proposal in the EBPXP model, which uses state-expected empowerment as a PX predictor. As a first step towards a proof-of-concept, we have explored which experiences EBPXP can potentially predict through a qualitative study on the custom-made game RoboRunner. Our findings support our research question RQ.9, ‘Can we use IR to predict people’s experience of interactive artefacts in a general and autonomous way?’, although only tentatively. Here, we discuss next steps to investigate our research question further and to enable the practical use of our approach (Sec. 8.3.1). We then highlight improvements to the underlying model, and discuss its potential application in two PCG scenarios (Sec. 8.3.2). Finally, we consider future work on applying our approach beyond games to other areas of CC, informed by the present findings (Sec. 8.3.3).

8.3.1 Directions for Further Study

Based on our qualitative study (Sec. 7.5), we have put forward the working hypothesis (Sec. 7.6) that state-expected empowerment does not directly predict goal experiences (Cairns, 2016) such as challenge, but the founda-
8.3 INTRINSIC REWARD-BASED EXPERIENCE PREDICTION

8.3.1 Intrinsic Reward-Based Experience Prediction

Figure 8.3: PX prediction on the GVGAI (Pérez-Liébana et al., 2019) game Frogs via the EBPXP model. 1-step state-expected empowerment has been calculated in our work-in-progress framework for all possible player positions at the current time step. Dark values indicate lower empowerment.

...tional experiences (Power et al., 2019) of effectance, outcome uncertainty and perceived control. A crucial next step to supporting our research question, and a prerequisite to the practical application of EBPXP is to evaluate this hypothesis through quantitative studies. Since the type of AI agent used to simulate human gameplay can severely impede the prediction accuracy, we propose to drop this component of our model and investigate the correlation of the IR and PX in an isolated fashion. This is possible by calculating IR post-hoc on the play trajectory of human players. We suggest employing both subjective and objective measures: while the first may be easier to obtain, the latter is often time-sensitive, and can thus be directly correlated with the IR along a recorded play trajectory. We propose to use questionnaires to measure the foundational experiences of outcome uncertainty (ibid.) and effectance (Klimmt, Hartmann & Frey, 2007), and to correlate them with the found goal experiences, in particular challenge (Denisova et al., 2020). The games user research literature could help in identifying appropriate objective measures for these experiences. At this point, we deem it interesting to correlate empowerment with a player’s electrodermal activity as a measure of arousal, and to employ electroencephalography to assess their attention (Nacke, 2013). The next step should be to conduct these quantitative studies on different variants of empowerment, as discussed below in Sec. 8.3.2. Our vision is to compile a mapping of IRs to the PXs they are best suited to predict, accounting for moderating factors such as the genre or a player’s expertise.

We have conducted our exploratory study only on a single, relatively simple game. Hence, we cannot tell whether the identified relationships between empowerment and PX persist in other games. Another crucial next step is thus to investigate whether our approach generalises. We propose to study EBPXP quantitatively on games from different genres and with different complexity. Starting with extensions of RoboRunner, we suggest a move to linear but substantially more complex games that still afford experimental control such as Infinite Mario Bros. (Persson, 2010), advancing to non-linear games that are still discrete, e.g. as part of the GVGAI framework (Pérez-Liébana et al., 2019), and eventually considering non-linear, continuous games such as VizDoom (Kempka et al., 2016).
To further this agenda, we have developed an evaluation framework on top of GVGAI. It allows us to record, store and replay human or AI play trajectories and hence disentangle gameplay from the PX estimation. Moreover, it can sample state transitions in stochastic games, and be extended for the calculation of various IRs. Fig. 8.3 shows a player’s state-expected empowerment in the Frogger (Konami, 1981) clone Frogs, evaluated with our framework. While GVGAI comprises mostly action games, puzzles and hybrids, these show a large diversity of mechanics. The games have been deliberately designed to challenge general game-playing agents within the GVGAI competition, and are thus promising candidates to probe the generality of EBPXP. GVGAI games are written in the Video Game Description Language (Schaul, 2014), which enables the straightforward construction of specific test scenarios. We propose to study first whether correlations between empowerment as IR and PX persist across different games, as described in the previous paragraph. These studies can then be extended by experiments with simulated gameplay, leveraging the readily available, state-of-the-art general game-playing agents included in the framework.

8.3.2 Improvements to and Potential Applications of Our Approach

EBPXP (Sec. 7.4) could realise several powerful features of our more generic, informally described approach if we manage to relax the simplifying assumptions of fixed parameters and full observability. Firstly, an agent that updates the parameters of their models while playing a game would calculate an adaptive, rather than a static (Oudeyer & Kaplan, 2007) empowerment reward. This may allow us to capture how changes to a player’s epistemic uncertainty, in contrast to the aleatoric uncertainty in the game world (Costikyan, 2013; Chua et al., 2018; Power et al., 2019; and Appx. A), influence their experience over time. If our working hypothesis in Sec. 7.6 proves correct, we could leverage adaptive EBPXP to estimate how foundational experiences such as outcome uncertainty (ibid.) and effectance (Klimmt, Hartmann & Frey, 2007) change as more gameplay experience is turned into better models. Linking these to goal experiences such as cognitive challenge (Cox et al., 2012; Denisova, Guckelsberger & Zendle, 2017) could enable intriguing applications such as the procedural generation of levels that remain challenging to the player. Secondly, dropping the requirement of full observability could facilitate the application of EBPXP to games with incomplete or imperfect information (e.g. Bard et al., 2020). Relaxing these assumptions requires us to replace the simplified vanilla empowerment (Eq. 3.14) in the state-expected empowerment (Eq. 7.1) as PX predictor with the general version from Eq. 3.9. As a first step towards this goal, we discuss exact and approximate Bayesian inference schemes for the model parameters θ in (Biehl et al., 2018).

The prediction of PX via IR (Sec. 7.3) relies on the simulation of game state trajectories via an accurate and general model of human player behaviour. We have highlighted in Sec. 7.6 that achieving such generality and human-likeness, especially in conjunction, is an open problem. We understand this as an opportunity for a mutually beneficial, iterative refinement process: while the PX assessment through simulated gameplay benefits from us making...
AI agents more human-like, the intrinsic reward-based PX modelling might provide insights on how players experience and consequently act in games, which could contribute to developing more general and human-like AI players.

We advocate three avenues for future inquiry. Firstly, we propose to motivate general game-playing agents intrinsically (Sec. 2.2) to overcome their reliance on human-defined extrinsic rewards, hence becoming more general. This could be realised by turning e.g. EBPXP into an on-policy approach (Sec. 7.3) in which the simulated agent realises EM\(^1\). Likewise, it could be realised in an off-policy manner where the agent optimises a (combination of) IRs different from the one used in the PX prediction. Procedural personas represent a particularly promising generative, off-policy model of archetypical player behaviour. Defined in terms of a utility vector describing player preferences, these models have been successfully used to reproduce human game-playing behaviour (Holmgård et al., 2014a, 2014b). To increase prediction accuracy, we secondly propose to align these simulations more closely with the style of a particular player by tuning the IM hyperparameters based on live player data. We expect the untuned agent to deliver a useful first approximation of human play, and to become better over time. Such an agent could thus also be used offline before live data becomes available. If this is not a requirement, a third avenue of inquiry becomes relevant: a player simulation model could be obtained through supervised learning on human play, or by reconstructing the human reward function from empirical playtraces via inverse reinforcement learning (Sutton & Barto, 2018, p. 470). Transfer learning promises to facilitate the application of human-like models of play acquired on one version of a game to another, and recent work in general game-playing (e.g. Pathak et al., 2017b) has shown that intrinsic reward can support such transfer.

We plan to instantiate our generic approach in Sec. 7.3 with other IR functions. We specifically propose to contrast the state-expected empowerment reward in EBPXP with a player’s empowerment dispersion as the standard deviation of their empowerment in all possible next game states. A value of zero means that no action is expected to yield a higher future potential and perceivable influence than any other. In contrast to state-expected empowerment, this variant thus relies on relative differences in empowerment rather than absolute values and has an arguably more intuitive baseline. We hypothesise that a player’s empowerment dispersion could predict experiences such as tension and decision-making challenge (Denisova et al., 2020).

We also strongly advocate substituting empowerment with other IRs (Sec. 2.2.4) and evaluating their capacity to predict PXs. One promising candidate is surprise, e.g. formalised as prediction error (Schmidhuber, 1991a), as it has been previously linked to the human experiences of boredom and surprise but also to people’s aesthetic judgement (Berlyne, 1971; Williams, 1996). We would be interested to see this reward being used in the prediction of PX arising from decorative game content (Smith, 2014b) such as graphics and sound to produce e.g. a less boring No Man’s Sky (Hello Games, 2016; Martin, 2016), but it could be equally applied to functional content.

Applications

We finally promote two potential applications of intrinsic-reward driven PX prediction in PCG that make extensive use of its putative flexibility and

\(^1\) The use of EM for general game-playing is supported by the related work in Sec. 5.2.1.
We firstly propose to employ our approach to take Smith and Whitehead’s (2010) concept of expressive range analysis further and calculate the experiential range of a game content space. For this instance of state-space characterization (Nelson, 2011), we would calculate IR-based PX predictions on a representative game content sample to visualise which degrees of a specific experience the game can realise. Embedded in a tool such as Danesh (Cook, Gow & Colton, 2016), this could be used to support game designers offline in probing the limitations of a content generator for a specific parameter configuration. By exploring the upper and lower bound on the possible mean state-expected empowerment of RoboRunner levels through evolutionary search in Sec. 7.5, we have already realised what Nelson (2011) considers a threshold strategy, contributing towards the goal of a full-on experiential range analysis by approximating the experiential boundaries of the possibility space. In contrast to existing tools that shed light on the relationship between a generator’s parameters and direct qualities of its output (Cook et al., 2019), the proposed experiential range analysis would show the expected impact of the PCG parameters on how the generated content will be experienced by players.

We would also like to see our approach used in the demanding task of automated game design (e.g. Cook, Colton & Gow, 2016a, 2016b). Liapis, Yannakakis and Togelius (2014) note that ‘autonomous computational game creators should attempt to design new games that can be both useful (playable) and deemed to be creative (or novel) considering that artifacts generated can be experienced and possibly altered’ (ibid., p. 46, emphasis added). We believe that our approach could evaluate subjective experience as part of a game’s value, and remain flexible enough to warrant the production of a large variety of different games without human involvement.

### 8.3.3 Next Steps in Advancing Computational Creativity

We set out to overcome major shortcomings of existing approaches to PX modelling in PCG, most notably a lack of flexibility to changes in the content generator caused by a strong dependency on people. In Sec. 7.1, we have identified a similar inflexibility in present techniques to evaluating the human subjective experience of interactive artefacts more generally. Here, we highlight central challenges in applying this approach to other types of artefacts and thus to benefit CC beyond computational game creativity (ibid.).

A player’s experience of a videogame as autotelic activity (Salen & Zimmerman, 2004, pp. 332-333) is largely shaped by events within the boundaries of its magic circle (cf. Huizinga, 1950, p. 10; and Sec. 5.1.1). Many other interactive artefacts however do not establish such a boundary, but may be explicitly designed to connect with our wider personal, social, cultural and political environment. To apply our approach to the evaluation of a specific interactive artefact, we must understand to which extent these contextual factors moderate the relationship between the audience’s IR in interacting with the artefact, and their experience. If the contextual moderation is too big, it may render our approach inapplicable. The reason is that we typically have little information about e.g. the context in which an artefact will be presented,
or the personal background of the audience members, and can hence not account for these factors in an a priori evaluation process.

We have highlighted the development of general models of human-like gameplay as an ongoing research challenge. Even fewer resources have been invested into, and considerably less progress has been made on, _modelling the interaction of an audience_ with other types of interactive artefacts. We suspect that this comes with particularly touch challenges in artistic domains, as e.g. an interactive film typically does not convey a specific goal towards which an audience interaction model could be oriented. For this reason, we propose to employ models of IM for the simulation of audience interaction. We deem models of _curiosity_ a good starting point, as they have been proposed for, and successfully employed in simulating the human aesthetic judgement of static artefacts (e.g. Macedo and Cardoso, 2001; Schmidhuber, 2006; Saunders, 2009; and Sec. 4.2.2). Especially for non-artistic domains, further inspiration may also be found in research on computational design (Saunders & Gero, 2004).

We can potentially mitigate the previous challenges and apply our approach more widely by reconsidering the concept of interactive artefacts. Our definition in Sec. 7.1 requires such an artefact to _change_ on user _interaction_. Crucially, this only distinguishes e.g. a kinetic sculpture from a traditional, ‘static’ one as long as we ignore that human perception is subjective and embodied. In reality, we do not directly access, but _model_ artefacts through sensorimotor interaction with our environment: we build a model of a sculpture by walking around and perceiving it, and a model of a poem by directing our gaze from line to line. This model undergoes constant change in response to our interaction. If we are ready to adopt this stance and consider a person’s reception process as one of interaction, we can apply our approach to a much wider range of artefacts for which reception models may exist, and which we may experience more independently of contextual factors.

### 8.4 Consolidation

We have considered many directions for developing the contributions of this thesis further over the short and long term, but always in isolation. We conclude this chapter with two examples of how they could be consolidated in future work to generate new insights within and beyond game AI.

In Sec. 8.2.2, we proposed tuning the CEM hyperparameters based on objective measures, assessed on the simulated interaction of the CEM-driven NPC with a game-playing agent. Crucially, measures such as performance difference only provide us with a rough idea of how a specific CEM parametrisation would impact a human player’s experience; especially in games with few opportunities to score, measures such as relative performance cannot tell us _when_ the NPC is challenging or supporting the player and _how_ strongly.

We propose to leverage EBPXP specifically and _intrinsic reward-based player experience prediction_ more generally to tune the parameters of a CEM-driven NPC towards a certain PX. Once validated, such models could allow us to track PX with high fidelity along the entire gameplay trajectory, and thus make sure that e.g. a companion NPC supports the player homogeneously over the course of the entire level, rather than just in the beginning.
Our second example addresses one of the foundations of this thesis: that human gameplay is shaped by the optimisation of IR, and can be described with computational models of IM. We have selected empowerment as IR for CEM because we expect that many instances of human play can be approximately described as an empowerment maximising process, and that increasing their empowerment would thus contribute to their progress in a game. Closely related, we also hypothesised that a player would experience such an increase in their empowerment as support, and a decrease as challenge. Future quantitative studies on EBPXP, as outlined in Sec. 8.3.1, will provide additional evidence to support or refute both these hypotheses: once we have recorded a person’s play and experiences by subjective and objective means, we can calculate empowerment on each state along their play trajectory. We can then validate whether they indeed strived to maximise the reward, and experienced areas of low or high empowerment as different degrees of challenge. This can be repeated for other types of IR, and consequently inform the development of alternative social models of IM that trigger certain PXs in the interaction with an NPC. Crucially, as games can simulate any facet of our reality at almost arbitrary levels of detail, these studies also have the potential to reveal much more far-reaching insights into human cognition: what ultimately motivates us as people.

In this chapter, we peeked into the future of IR and IM models in game AI and CC through the lens of our own theoretical and applied work. This marks our last contribution, and we conclude this thesis in the next chapter.
CONCLUSION

We motivated this thesis with a crucial observation: despite the important role of intrinsic motivation (IM) in human cognition and creativity (Ryan & Deci, 2000a; Amabile, 2018), the majority of computational creativity (CC, Colton and Wiggins, 2012) systems are extrinsically motivated. We argued that this motivational focus counteracts core CC research goals, in that it can negatively impact people’s assessment of creativity (Colton, 2008) in these systems, their actual creativity, and their creative autonomy (Jennings, 2010). A small body of prior work has explored the alternative route of building CC systems based on computational intrinsic reward (IR) and models of IM (Oudeyer & Kaplan, 2007). However, this work focusses on specific motivational mechanisms, unaware of the possibility of understanding them as instances of a bigger family, which as a whole may hold benefits for CC. Our aim for this thesis has consequently been to investigate whether computational IR and models of IM more generally, as a distinct class of motivation and as a family of mechanisms, can advance central goals of CC. We approached this challenge through the lens of computational game creativity as ‘the study of computational creativity within and for computer games’ (Liapis, Yannakakis & Togelius, 2014, p. 2). More specifically, we have expanded our research aim to game AI to synergistically further insights on the advantages of IM in both domains.

We adopted a highly interdisciplinary approach and multiple perspectives to assemble a big picture view of the benefits and applications of IR and IM models in CC and game AI. As a foundation to this account, we have synthesised insights from psychology and AI into a working definition of IM models (Ch. 2). We then informed this big picture by considering the past, present and future of employing IR and IM in CC and game AI. We mapped research by means of two systematic reviews of related work in both domains (Ch. 4 and 5), drawing on insights from creativity studies and (computational) creativity theory as well as game design and games user research, respectively.

We complemented this retrospective account by proposing, formalising and evaluating two new applications of IR and IM to game AI through the lens of CC. We hereby leveraged empowerment and EM (Ch. 3) as specific IR and IM model. We firstly (Ch. 6) proposed social models of IM as a means to drive the behaviour of general, co-creative artificial agents that can support or challenge their partner in an open-ended way. We instantiated this proposal in coupled empowerment maximisation (CEM) to drive the behaviour of general and believable non-player characters (NPCs) that either support or challenge a player as companions and adversaries, respectively. We defined and employed the qualitative method of observational vignettes to evaluate the behaviour of CEM-driven NPCs in two studies. Secondly (Ch. 7), we proposed to evaluate people’s experience of interactive artefacts based on IR, as a means to make such evaluation in CC less dependent on people and a specific context. We instantiated our informal proposal in empowerment-based player experience prediction (EBPXP) for the application in videogame
procedural content generation (PCG). We employed an experiential vignette to explore qualitatively which experiences empowerment could predict. Our two applications allowed us to illustrate the advantages of IM in CC and game AI from very different perspectives: the (i) use of IM models vs. formal IR in the (ii) CC generation vs. evaluation of (iii) simple creative behaviour vs. complex artefacts, applied to (iv) NPCs vs. PCG as game AI domains.

We completed our big picture view by highlighting promising future directions for the exploration of IR and IM models in CC and game AI based on improvements of our own studies, the underlying approaches, and potential applications to other CC domains and beyond (Ch. 8).

We next summarise key findings for our specific research questions, and discuss their contribution to our overarching research aim (Sec. 9.1). We then reiterate our contributions and discuss their potential impact on various areas of academic and industrial research (Sec. 9.2). We end this thesis with a few concluding remarks (Sec. 9.3).

9.1 Research Questions Revisited

In the introduction to this thesis (Ch. 1), we formulated our research aim in terms of two overarching research questions:

RQ.1 Can IR and models of IM advance CC?

RQ.2 Can IR and models of IM advance videogame AI?

We qualified these research questions with seven specific research questions, and evaluated them through five qualitative studies: two systematic reviews, two observational vignette studies on CEM-driven NPCs, and one experiential vignette study on empowerment-based PX prediction. We highlight selected findings for each specific question, and their contribution to RQ.1 and RQ.2.

We affirm RQ.1 through these two retrospective questions posed in Ch. 4:

RQ.3 Why have IR and models of IM been used in CC?

RQ.4 How have IR and models of IM been used in CC?

Our answers rest on a systematic review of 29 theoretical and applied studies in CC, dating from 1998 to 2018. Our classification in Fig. 4.1 reveals that these studies are well-balanced across different application domains and system types (e.g. autonomous/co-creative, single/multi-agent, etc.), which partially answers RQ.4. Moreover, we have identified four properties of IR, two corollaries, and four properties of intrinsically motivated behaviour as reasons to embrace IM in CC and as an answer to RQ.3. We have also extracted 12 (abstract) applications of IM to CC, which complements our prior answer of RQ.4. Our full answer is given by the two typologies in Fig. 4.2. Here, we only describe two diverse applications which leverage properties of IR and intrinsically motivated behaviour, respectively.

IR has been used e.g. by Macedo and Cardoso (2001b) to model p-creativity (Boden, 1990/2003, p. 1, 43 ff.). This application leverages two emergent properties of IR. A specific IR function can be defined with or be attributed different semantics (R.1 in Fig. 4.2), and can thus be used to quantify novelty...
and value (C.1 in Fig. 4.2) as essential components of creativity (Runco & Jaeger, 2012). Moreover, IR is subjective and sensitive to an agent’s embodiment and situatedness (R.2 in Fig. 4.2). Together, these properties enable the calculation of creativity from an agent’s own perspective. Moreover, intrinsically motivated behaviour has been used e.g. by Maher, Merrick and Macindoe (2005) to model forms of mini-e (Kaufman & Beghetto, 2009) acts (A.12 in Fig. 4.2), based on the potential of IM to induce skill and model development (B.4 in Fig. 4.2), and to yield open-ended adaptation (B.3 in Fig. 4.2).

We analogously affirm RQ.2, i.e. the same question directed to game AI, through the following two retrospective questions posed in Ch. 5:

RQ.5 Why have IR and models of IM been used in videogame AI?

RQ.6 How have IR and models of IM been used in videogame AI?

Our literature search revealed that IM is at present mostly applied to drive general game-playing agents as a benchmark for artificial general intelligence; only a few contributions aim to benefit game engineers, designers and players. To further draw a diverse, representative picture of the reasons to embrace IR and IM models in game AI and their application, we biased our review towards applications in the latter category, which yielded 11 instances of related work from 2006 to 2019. We classified this work in Tbl. 5.1 as partial answer to RQ.6. Amongst others, we find that many applications contribute towards several game AI domains at once, e.g. the design of game-playing agents and NPCs. Moreover, most of them aim to increase the generality of the respective game AI technique. Over all domains, PCG has been addressed the least, in only a single theoretical study (Shaker, 2016). As answer to RQ.5, we crucially uncover the same reasons to embrace computational IR and models of IM found in existing CC work (Ch. 4). We moreover identify 11 (abstract) applications of IR and IM across four core domains of videogame AI. We provide examples of two applications.

Shaker (ibid.) suggests the usage of intrinsically motivated RL to create content without game domain knowledge (A.9 in Fig. 5.1), e.g. when improvising ‘new types of games from scratch’ (ibid., p. 455). Her theoretical proposal rests on the domain and embodiment generality of IR (R.3 in Fig. 5.1). The same property is used by Pathak et al. (2017b) to enable transfer learning from one game or game level to another (A.1). This application also makes use of the fact that many IM models induce skill and model development (B.4 in Fig. 5.1). They demonstrate that a game-playing agent can use IM to facilitate transfer learning across different levels of the same game.

In an effort to support RQ.1 through insights in RQ.2 and vice versa, we answer the following question in Ch. 5 through our two systematic reviews:

RQ.7 How do existing applications of IR and IM models in videogame AI and CC overlap?

We find that that many applications of IR and IM in game AI specialise CC applications. For instance, we can understand the efficient exploration of game content spaces (A.10 in Fig. 5.1) to produce novel and valuable content in videogame PCG as a special case of exploratory and transformational creativity (Boden, 2003; Wiggins, 2006; A.6 and A.8 in Fig. 4.2) in CC. Similarly,
we recognise applications of IR and IM models to (game-)playing and NPCs as specialisations of CC applications. These game AI applications leverage e.g. the potential of IM to respond autonomously to unanticipated events or changes in complex, open-ended game worlds (A.8 in Fig. 5.1), which echoes the CC application to model mini-c creativity in development and adaptation (A.12 in Fig. 4.2) and to increase creative autonomy (A.3 in Fig. 4.2). We highlight this example specifically because (game-)playing and NPCs have so far received little attention in discussions of computational game creativity.

Moreover, we find that applications in game AI not only specialise existing CC applications, but also advance CC by realising different notions of generality. This property is rarely addressed in mainstream CC research, despite necessitating different forms of creativity. For instance, for the same game-playing agent to perform well across different games, the agent must express both novelty and value, and thus creativity, in their behaviour. These insights thus reveal that individual answers to the questions RQ.1 and RQ.2 also support their counterpart.

We further support that IR and IM models can be leveraged for both, CC and game AI through two novel computational approaches, motivated in both domains and evaluated in game AI. These address underexplored areas of research identified in our systematic reviews, but also draw inspiration from existing CC (Saunders & Gero, 2004; Saunders et al., 2010) and game AI (Togelius & Schmidhuber, 2008; Merrick & Maher, 2009) research.

In Ch. 6, we address RQ.1 and RQ.2 through the following question:

**RQ.8** Can we use a model of intrinsic motivation to engineer general and social co-creative agents?

We answer this question by proposing *social models of IM* as a means to overcome the shortcomings of related work, instantiating this proposal in CEM, and evaluating the potential of this computational model to drive the behaviour of general, believable NPCs that support or challenge the player as companions or adversaries, respectively. This link rests on our understanding of the player-NPC interaction as an instance of human-computer co-creativity.

Our evaluation rests on 14 experiments as part of six observational vignettes, split between two exploratory, qualitative studies. They demonstrate that CEM as a single principle can indeed be used to yield both supportive and antagonistic behaviour by only switching a single hyperparameter. Moreover, we discovered that the modulation of the hyperparameters affords the creation of different NPC personalities such as a daredevil and a super-villain. As predicted, the CEM-driven NPCs exhibited *player-generality* (Togelius & Yannakakis, 2016) in consistently supporting or challenging the different experimenters as players without modelling their specific policy. Moreover, they demonstrated limited *game-generality* (ibid.) by maintaining these social dynamics in response to changes in their embodiment and environment.

These behaviours have not been hard-coded, but emerged from the interaction of CEM, the NPC’s embodiment, environment and their interaction partners. The emergent behaviours were thus novel, and often took us by surprise. Moreover, they have value with respect to a person’s goals in three forms of co-creativity: in supporting the player’s game goals as companions, in providing them with the experience of interacting with a specific character,
9.2 Contributions and Potential Impact

We have previously discussed how our theoretical and applied contributions have supported the overarching aim of this thesis. We next highlight their potential future impact beyond this thesis on AI, CC and game AI research, as well as on game engineers, designers and players.

In Ch. 2, we have contributed an extensive, interdisciplinary account of IM in psychology and AI (Sec. 2.1). Moreover, we developed and tested an informal working definition of IM models, based on four diagnostics of a reward function embedded in a motivational model (Sec. 2.2.3). These contributions have served as a foundation to our big picture view of IM in CC and game AI, but we believe that they can also benefit other disciplines. Our overview can allow psychologists and AI researchers to familiarise themselves with the treatment of IM in the other domain and inspire future cross-disciplinary research. We expect our close comparison of the diagnostics of formal IR and IM and the psychological definition of IM in Sec. 2.2.3 to

RQ.9 Can we use IR to predict people’s experience of interactive artefacts in a general and autonomous way?

We approach this question with an informal and generic proposal to predicting people’s experience via IR. We instantiate this proposal in EBPXP and assess it on the prediction of players’ experiences of procedurally generated content. We hence answer this question through an application to game AI, by understanding a procedural content generator as a CC system.

We have conducted an exploratory, qualitative study to inform a future quantitative study with candidate experiences that EBPXP could predict. We chose the method of an experiential vignette, and asked our participants to think-aloud while playing different game levels, each procedurally generated as a distinct condition expressing a different EBPXP prediction. A thematic analysis of the think-aloud data revealed that our conditions indeed evoked different experiences. We found the most striking differences in players’ experience of physical and cognitive challenge, followed by their involvement, attention, engagement and emotions. Following a critical evaluation of our findings based on games user research, we deem most of these experiences too complex to be directly predicted by empowerment as IR. By considering commonalities between these experiences, we develop the hypothesis that empowerment can predict the foundational experiences of effectance, outcome uncertainty, and perceived control, which influence goal experiences such as challenge. While our study provides some evidence to support RQ.9, a full proof-of-concept requires further work, as discussed in Ch. 8.

Together, our answers affirm the overall research questions RQ.1 and RQ.2 through a big picture of the benefits of IR and IM models for CC and game AI, and their application in past, present and future research.
be particularly helpful in this respect. We moreover hope that our working definition will inspire future discussions on the nature of IM, and eventually lead to a formal definition of IR and IM models.

In Ch. 3, we contributed an updated introduction to empowerment and EM in discrete scenarios, with the strongest psychological motivation available to date (Sec. 3.1). Moreover, we provided a generic and simplified formalisation of empowerment and EM, which distinguishes an agent’s objective world and their beliefs about that world, and thus makes implicit assumptions in prior work transparent (Sec. 3.2). There has recently been a surge of interest in empowerment and EM specifically, in particular employed in an RL setting to further the goal of artificial general intelligence. We hope that our comprehensive motivation of EM will further illuminate its potential towards this end and inspire new research, e.g. in collaboration with cognitive scientists. Similarly, we hope that our generic formalisation will raise people’s awareness of the differences between an objective and epistemic account of empowerment, which could again shape future research.

In Ch. 4 and 5 we presented two systematic reviews of existing work on IM in CC and game AI as the key theoretical contributions of this thesis. These comprise a classification of existing work in both domains (Tbl. 4.1 and 5.1). We moreover contributed typologies on the reasons to embrace IR and IM models, and their individual application in both domains (Fig. 4.2 and 5.1). We hope that these typologies will allow CC and game AI researchers to identify promising areas of future inquiry, and serve as inspiration and reference to harness the benefits of IR and IM in their work. We moreover hope that our mapping between these applications in CC and game AI have illustrated how strongly these fields can complement each other, and thus prompt future research in computational game creativity.

In Ch. 6, we contributed a new application of IM to CC and game AI in the form of social models of IM as a generic approach to yield supportive or adversarial agent behaviour in open-ended interaction (Sec. 6.3). Moreover, we introduced CEM as a specific social IM model to give rise to general, believable companion and adversary NPCs (Sec. 6.4). CEM could benefit game engineers, designers and players. It could increase the believability of NPCs and thus heighten how players experience and enjoy the next generation of videogames. At the same time, it could speed up NPC development and decrease production costs, thus allowing a reallocation of engineering resources. It could further afford more creative freedom to game designers, by enabling more complex and dynamic game worlds, the integration of user-authored content, and more sophisticated PCG. CEM could in addition serve as a source of design inspiration, in that CEM-driven NPCs exhibit behaviours that are novel and surprising. They could also transform the NPC design practice: rather than hard-coding a specific character for a game world, designers could take the opposite position and observe what emerging behaviours, and thus characters, a specific game world gives rise to.

Kantosalo and Toivonen (2016) predict that ‘in the future, systems taking a more provoking stance may be of particular interest for co-creativity research (...)’ (ibid., p. 83, emphasis added). CEM represents a candidate model to realise such systems in wider CC, but we also promote its use beyond human-computer co-creativity. For instance, CEM could be utilised in guiding the
emerging social dynamics of creative agent societies (Saunders & Bown, 2015) to yield a particular, human-desired output. We furthermore advocate the use of CEM beyond traditional CC scenarios whenever embodied agents are required to support or challenge other agents in complex, changing environments and for different embodiments. We envision it to be used in robot-robot interaction, e.g. to allow a swarm of heterogeneous rescue robots to reconfigure flexibly in order to overcome debris, or to operate more autonomously in remote locations. Moreover, Salge and Polani (2017) propose using CEM to warrant safety, compliance, and robustness in human-robot interaction without the need for natural language understanding.

In Ch. 7, we introduced a new application of IR to CC and game AI. Motivated by challenges in evaluating people’s experience of interactive artefacts in CC more generally, we proposed intrinsic reward-based player experience prediction as a generic approach to modelling player experience independently of player feedback and designer knowledge about a game’s semantics. As a second contribution, we instantiated this approach in EBPXP and evaluated it on the prediction of players’ experience of procedurally generated game content. This complements our previous applied contribution in that we do not use models of IM to steer behaviour, but leverage the underlying reward as an experience predictor.

Given a successful proof-of-concept, our approach could yield improved experiential control of procedurally generated game content, as changes to the game or content generator would not necessitate the costly and time-intensive provision of player data or changes to the theoretical model assumptions. This would enable the fast and frequent assessment of PX within the development process, and for widening the expressive range (Smith & Whitehead, 2010) of generators employed online during play. The freed resources and increased creative potential could inspire designers to develop novel games with improved replayability, potentially leading to higher player satisfaction. Without the need to continuously involve human players and designers, our approach could even be leveraged in the demanding scenario of automated game design (e.g. Cook, Colton & Gow, 2016a, 2016b).

Beyond game AI, our approach could complement established instruments to measure PX in games user research. Applied ahead of a user study, predictions of which PXs an experimental condition is expected to cause could be used to adjust the condition to the study goals, or to identify the best subjective or objective measures to accurately capture the predicted experiences.

Finally, we believe that our generic approach could be employed in the evaluation of user experience more generally, as supported by the work of Trendafilov and Murray-Smith (2013). We hope that the findings of our experiential vignette studies encourage researchers to examine the relationship between other models of IR and user experience.

9.3 Concluding Remarks

In this thesis, we have demonstrated that, and explored why and how IR and models of IM can advance CC and game AI, by drawing on pioneering work, proposing new applications, and envisioning the future. This endeavour has
substantially benefited from a highly interdisciplinary perspective, and from leveraging the symbiotic relationship between CC and videogames. Writing this thesis has been a thought-provoking and rewarding experience, and we hope that our approach and findings will inspire researchers to pursue a similar path in the future and advance knowledge in both domains.

We are excited to see how our applications will be adopted and developed further to increase the generality, autonomy and creativity of next-generation videogame AI, and consequently transform research and industry practice as well as the experience of players. We moreover hope that this thesis will increase awareness of the IM concept, contribute to a better understanding across disciplines, and serve as inspiration for future work. We are confident that comprehending the possibilities of IM moves us closer to tackling one of the major challenges of CC: to engineer artificial systems that are creative.
In this interdisciplinary thesis, we investigate computational models of intrinsic reward and intrinsic motivation to drive the interaction of an agent with their environment. The following appendices summarise the mathematical foundations required to formalise each of the involved concepts, and clarify our notation. We only assume little prior knowledge to make our work accessible for researchers from a wide range of related disciplines.
The interaction of an embodied agent with their environment is subject to different forms of uncertainty. **Probability theory** allows us to make formal statements about uncertainty, and we thus summarise some fundamental probabilistic concepts here. This primer largely draws on the engineering-focussed introduction by Bishop (2006). Uncertainty can originate from different sources (cf. Jaynes, 2003): Firstly, uncertainty can be inherent to the system being modelled (**aleatoric uncertainty**), e.g. when a random number generator is being used to determine the damage of a weapon in a videogame. Secondly, uncertainty can stem from an agent’s limited capacity to observe that system (**epistemic uncertainty**), e.g. when subjected to hidden information or partial observability: in a game of two players, the controlled character may have no information on the other player’s policy, or can only perceive the other’s actions but not their remaining health. Finally, an agent might be capable of observing a system’s state perfectly, but discard parts of the information and thus create uncertainty where there none (**self-inflicted uncertainty**). A game character could measure the game’s state precisely but discard some of the contained information, e.g. regarding its previous actions. Such ‘forgetting’ is often enforced by bounded rationality (Simon, 1957), e.g. based on memory or information processing limitations.

**Random variables** allow us to capture such uncertain aspects of the system; they can take on different **values (or states, or realisations)** based on an underlying random process. We denote random variables by upper case letters, e.g. $X, Y, S, A, R$, etc., and specific values of these variables by lower case letters $x, y, s, a, r$, etc. Random variables can take on a finite number of **discrete** values, or an infinite number of **continuous** values. We denote these possible values, i.e. the **state space** or **range** of a random variable, by the corresponding calligraphic upper case letters $\mathcal{X}, \mathcal{Y}, \mathcal{S}, \mathcal{A}, \mathcal{R}$, etc.

**Probability distributions** allow us to describe how likely a random variable is to take on its possible values. In the case of discrete random variables, a probability distribution is defined by a **probability mass function**, mapping the values of a random variable to the probability that the variable takes on these values, i.e. $p : x \mapsto [0, 1]$. When a random variable $X$ is distributed according to $p$, we write $X \sim p$. A probability mass function over a random variable $X$ must satisfy three properties. Firstly, its domain must contain all possible states $x \in \mathcal{X}$ of $X$. Furthermore, it must satisfy

$$\forall x \in \mathcal{X} : 0 \leq p(X = x) \leq 1 \quad \text{(A.1)}$$

$$\sum_{x \in \mathcal{X}} p(X = x) = 1 \quad \text{(A.2)}$$

i.e. all states must be assigned values between 0 (impossible) and 1 (guaranteed to happen), and the probabilities of individual states must sum up to 1.
When a probability mass function describes the probabilities of more than one random variable, e.g. \( p(X = x, Y = y) \), we refer to a joint distribution. Usually, the state of a random variable is denoted by the same letter as the variable. When relating to the probability of a specific state, we often omit the corresponding random variable, so \( p(X = x) \) and \( p(X = x, Y = y) \) become \( p(x) \) and \( p(x, y) \), respectively. Similarly, we usually omit the state space in summation for brevity (Cf. Eqs. A.2 vs. A.17).

For continuous random variables, a probability distribution can be defined by means of a probability density function \( f(X = x) \), specifying the relative likelihood that the random variable \( X \) would take on value \( x \), i.e. \( f : x \mapsto \mathbb{R}_{\geq 0} \).

A probability density function must be non-negative everywhere, and its integral over the entire space must be equal to one:

\[
\int_{-\infty}^{\infty} f(X = x) \, dx = 1 \tag{A.3}
\]

The probability that a continuous random variable \( X \) takes on values no larger than \( x \) is given by the cumulative distribution function:

\[
p(X \leq x) = F(X = x) = \int_{-\infty}^{x} f(X = t) \, dt \tag{A.4}
\]

The probability that \( X \) takes values in the semi-closed interval \((a, b]\) is:

\[
p(a < X \leq b) = F(X = b) - F(X = a) \tag{A.5}
\]

In this thesis we make frequent use of two very common, discrete probability distributions, the discrete uniform distribution and the categorical (multinoulli) distribution. We use them as examples to illustrate the above concept of a probability mass function. If a random variable \( X \) is uniformly distributed, i.e. \( x \sim \mathcal{U} \), all values \( x \in \mathcal{X} \) are assumed to have equal probability:

\[
p(x) = \begin{cases} 
  \frac{1}{|\mathcal{X}|} & \text{if } x \in \mathcal{X} \\
  0 & \text{otherwise.} 
\end{cases} \tag{A.6}
\]

Here, \(|\mathcal{X}|\) represents the cardinality of the state space, i.e. the number of values it contains. This is a proper probability mass function in that it covers all possible values and satisfies Eqs. A.1 and A.2.

By default, we assume random variables in this thesis to be categorically (multinoulli) distributed. The probability mass function of a categorical distribution is defined, i.e. parametrised, by a vector \( \phi = [\phi_1, \ldots, \phi_{|\mathcal{X}|}] \) with \( \phi_i \) representing the probability of the \( i \)th value \( x_i \), i.e. \( p(x_i) = \phi_i \). The categorical distribution can describe any possible distribution over its domain accurately, at the cost of an inefficient description which enumerates the probability of each state. This however fits the purpose of the experiments presented later, where we are interested in exhaustive proof-of-concepts rather than approximations, and where state spaces are reasonably small. For brevity, we usually omit the parameter \( \phi \) from the description of categorical distributions and only refer to it explicitly when describing its estimation in pseudocode.
We sometimes leave the specific type of distribution unspecified. We then include parameters explicitly and write them as upper case Greek letters \( \Phi, \Theta \), etc. with values \( \phi, \theta \), etc. In our notation, an arbitrary parametrised distribution on \( X \) is then written as \( p(X; \Theta) \).

A special case is given if the distribution of interest, e.g. \( p(X) \), only takes on a single value of its random variable \( x \) with absolute certainty. If this variable is discrete, we specify such distributions using the Kronecker delta:

\[
\delta_{x_j, x_i} = \begin{cases} 
1 & \text{if } x_i = x_j, \\
0 & \text{otherwise.}
\end{cases}
\]  

The expression \( p(x) = \delta_{x, x_i} \forall x \in \mathcal{X} \) then describes a distribution which is 1 at value \( x_i \) and 0 for any other value in \( \mathcal{X} \). Summation over all \( x_i \in \mathcal{X} \) yields 1, and the requirements for a probability mass function are thus met. In few cases, we assume a similar distribution over continuous variables, and use the Dirac delta as shorthand. It is infinite at the origin, and zero everywhere else:

\[
\delta(x) = \begin{cases} 
\infty & \text{if } x = 0, \\
0 & \text{otherwise.}
\end{cases}
\]  

The Dirac delta satisfies the properties of probability density functions, i.e. it is non-negative and integrates to 1 (Eq. A.3). The notation is different from Kronecker’s delta: if we assume a distribution to take on the value \( x_i \) of a continuous random variable \( X \) with certainty, we write \( p(x) = \delta(x - x_i) \).

We make intensive use of conditional probabilities to describe the probability of some variable given the known state of another variable:

\[
p(x|y) = \frac{p(x, y)}{p(y)}
\]  

Here, \( p(y) > 0 \). Note that a conditional probability per se does not imply causality: \( x \) is not necessarily a consequence of \( y \). We introduce interventional distributions which represent such causality in Appx. B.

Two variables \( X, Y \) are (stochastically) independent, i.e. \( X \perp \! \! \! \! \perp Y \), if their joint distribution can be expressed as product of individual distributions:

\[
(X \perp \! \! \! \! \perp Y) \Leftrightarrow p(x, y) = p(x)p(y) \forall x \in \mathcal{X}, y \in \mathcal{Y}
\]  

This must hold for all, not only some values of the involved random variables. We can find an equivalent expression by applying Eq. A.9, the definition of conditional probability, to Eq. A.10:

\[
(X \perp \! \! \! \! \perp Y) \Leftrightarrow p(x|y) = p(x) \quad (A.11) \\
\Leftrightarrow p(y|x) = p(y) \forall x \in \mathcal{X}, y \in \mathcal{Y} 
\]  

We furthermore say that \( X \) is conditionally independent of \( Y \) given a third variable \( Z \), i.e. \( (X \perp \! \! \! \! \perp Y) \mid Z \), if the following condition holds:

\[
(X \perp \! \! \! \! \perp Y) \mid Z \Leftrightarrow p(x|y, z) = p(x|z) \forall x, y, z \in \mathcal{X}, \mathcal{Y}, \mathcal{Z}
\]  

\[\text{Independence}\]
In analogy to Eq. A.10, we can derive an equivalent condition, expressing that the joint distribution of \(X, Y\) is stochastically independent given \(Z\):

\[
(X \perp Y \mid Z) \Leftrightarrow p(x, y \mid z) = p(x \mid z)p(y \mid z) \quad \forall x, y, z \in X, Y, Z
\]

(A.14)

Even though they sound similar, the equivalent conditions expressed by Eqs. A.10–A.12 are different from the conditions in Eqs. A.13 and A.14.

Using the definition of conditional probability (Eq. A.9), we can write any joint distribution as a product of conditional probabilities. For two variables \(X, Y\), we get the product rule:

\[
p(x, y) = p(x \mid y)p(y) = p(y \mid x)p(x)
\]

(A.15)

By replacing one of the random variables with a set of variables and applying Eq. A.15 recursively, we can derive the chain rule or general product rule:

\[
p(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} p(x_i \mid x_1, \ldots, x_{i-1})
\]

(A.16)

Given a joint distribution, we can calculate the distribution over any subset of its members. The process of summing over the left-out variables is called marginalisation. A marginal distribution for discrete variables is given by:

\[
p(x) = \sum_y p(x, y)
\]

(A.17)

Note that the sum runs over all elements \(y \in Y\), but we have abbreviated this expression as announced earlier. The law of total probability allows us to marginalise over conditional probabilities, by applying Eq. A.15 to Eq. A.17:

\[
p(x) = \sum_y p(x \mid y)p(y)
\]

(A.18)

In the case of continuous variables, the sum in the calculation of the marginal distribution and total probability is replaced by an integral.

The final concept required in our probability theory toolbox is the expectation of a function \(f(x)\). It corresponds to the average value the function takes on if its arguments are drawn from distribution \(p(x)\):

\[
\mathbb{E}_{X \sim p}[f] = \sum_{x} p(x)f(x)
\]

(A.19)

For continuous random variables, the sum must be replaced by the integral over \(x\). In most cases the distribution of \(X\) can be inferred from context, and it might even be clear with respect to which random variable the expectation is taken. For brevity, we then write \(\mathbb{E}_X[f(x)]\) or just \(\mathbb{E}[f(x)]\), respectively.

In Appx. B, we introduce causal Bayesian networks as compact representations of joint probability distributions, and vice versa as a means to derive such joint distributions from a more accessible, graphical model.
CAUSAL BAYESIAN NETWORKS

Empowerment maximisation (EM) as the central model of intrinsic motivation (IM) investigated in this thesis requires us to consider the causal influence between different entities in the interaction of an agent with their environment. This can be conveniently done by means of causal Bayesian networks, which we introduce here. Causal Bayesian networks are a special kind of probabilistic graphical model. We briefly introduce such models and their benefits. We then define (non-causal) Bayesian networks and explain how non-local conditional independence assumptions can be identified in the network structure. This allows us to point out an ambiguity in the representation of causality in Bayesian networks (BNs), and motivates the introduction of specialised causal Bayesian networks. We mainly draw on the comprehensive introductions to Bayesian networks by Bishop (2006) as well as Koller and Friedman (2009), and on Pearl’s (2000) original work on causality.

A probabilistic model encompasses a set of random variables with assigned probability distributions, and is used to model a certain phenomenon. A probabilistic graphical model is a diagrammatic representation of a probabilistic model, i.e. it is defined as a graph \( G = (V, E) \) with vertices (or nodes) \( V = \{V_1, V_2, \ldots, V_m\} \) representing random variables and edges \( E = V \times V = \{e_1, e_2, \ldots, e_n\} \) describing probabilistic relationships between these variables. Crucially, a node can represent a single – potentially vector-valued – random variable, but also sets of variables. Graphical models complement probabilistic models in several ways (cf. Bishop, 2006, p. 360): They can (i) serve as a primary description or a secondary representation of a probabilistic model. As such, they (ii) afford the description and identification of model properties such as conditional independence. Finally, they (iii) allow us to express complex computations in terms of simpler graphical manipulations.

Bayesian networks (BNs) or directed graphical models are a specific class of probabilistic graphical models. A BN (Fig. B.1) is defined (cf. Koller & Friedman, 2009, p. 62) as a tuple \( B = (D, P_B) \). The first component \( D = (V, E) \) is a directed, acyclic graph. We briefly define these two qualifiers. As in general graphical models, nodes represent (sets of) random variables, but here bidirectional edges have been replaced by directed arrows \( V_i \rightarrow V_j \), representing the direct influence of variable \( V_i \) over \( V_j \). When there is an arrow from \( V_i \) to \( V_j \), we say that \( V_i \) is a parent of \( V_j \), i.e. \( \text{pa}(V_j) = \{V_i\} \). A path in \( B \) is a sequence of at least two vertices \( V_i \in V \) where adjacent vertices are connected by a single edge. In a directed path, all arrows must point in the same direction. We say that the graph is acyclic, when there is no closed path \( (V_i \rightarrow V_j \rightarrow \ldots \rightarrow V_i) \), i.e. there is no directed path starting and ending at the same vertex. This summarises the topology, i.e. the network’s first component. The second component \( P_B \) is a set of conditional probability distributions \( p_B(V_i|\text{pa}(V_i)) \), one for each node \( V_i \in V \), and each conditioned on the smallest set of predecessors \( \text{pa}(V_i) \) that renders \( V_i \) independent of all its other predecessors, as specified by the topology. Crucially though, \( V_i \) can
still depend on its descendants. The distributions in $P_B$ thus only encode the local independence assumptions $(V_i \perp nd(V_i) \mid pa(V_i))$, with $nd(V_i)$ being variables in $V$ that are not descendants of $V_i$.

By exploiting conditional independence assumptions, a BN can serve as a compact representation of a joint distribution. Consider factorising $p(v) = p(v_1, v_2, v_3, v_4, v_5)$ with the chain rule (Eq. A.16). For this specific ordering of variables in the joint distribution, we get:

$$p(v) = p(v_1)p(v_2 \mid v_1)p(v_3 \mid v_1, v_2)p(v_4 \mid v_1, v_2, v_3)p(v_5 \mid v_1, v_2, v_3, v_4)$$  \hspace{1cm} (B.1)

We can make this factorisation more compact by applying the local conditional independence assumptions encoded in the BN in Fig. B.1:

$$p(v) = p(v_1)p(v_2 \mid v_1)p(v_3 \mid v_1)p(v_4 \mid v_2, v_3)p(v_5 \mid v_3)$$  \hspace{1cm} (B.2)

Here, we have implicitly applied the chain rule for BNs:

$$p(v_1, v_2, \ldots, v_m) = \prod_{i=1}^{m} p_B(v_i \mid pa(v_i))$$  \hspace{1cm} (B.3)

If a distribution $p$ factorises over the topology $D$, it satisfies the local conditional independence assumptions in $D$. However, the topology encodes further, non-local conditional independence assumptions that $p$ must meet.

We can identify such further assumptions with the concept of d-separation (Pearl, 1988), the graph equivalent of probabilistic conditional independence. To begin, we define that a path is blocked if it contains a node $V_i$ and if either:

- $V_i \not\in S$ and the arrows meet at $V_i$ head-to-tail ($\rightarrow V_i \rightarrow$ or $\leftarrow V_i \leftarrow$) or tail-to-tail ($\leftarrow V_i \rightarrow$),

- or neither $V_i$ nor any of its descendants are in $S$, and arrows meet $V_i$ head-to-head ($\rightarrow V_i \rightarrow$).

A set of variables $A$ is then d-separated from a set $B$ by a third set $S$ with implication that $(A \perp B) \mid S$ if all paths from $A$ to $B$ are blocked by $S$. Consider the previous example network in Fig. B.1 for two illustrations of d-separation. Here, we have $(V_2 \perp V_3) \mid V_1$ because one path is blocked by $V_1$ which is met tail-to-tail and the other path is blocked by $V_4$ which is not part of the conditioning set and met head-to-head. As a negative example, we have $(V_1 \not\perp V_5) \mid V_2$ despite one path being blocked by $V_2$ and $V_4$, because the remaining path through $V_3$ is not blocked.
Applying the d-separation criterion to the three graph structures in Fig. B.2, we find that they express the same non-local conditional independence assumption, despite encoding different local independences. We consider such graphs I-equivalent (Koller & Friedman, 2009, p. 76) or observationally equivalent (Pearl, 2000, p. 19), as they cannot be distinguished based on observational data alone. Any distribution $p$ that can be factorised over one I-equivalent graph can also be factorised over other graphs in the same equivalence class; there is no inherent property of $p$ that warrants the association with one over another I-equivalent graph. A distribution $p(V_1, V_2)$ for instance can be factorised to both $p(V_2)p(V_1|V_2)$ and $p(V_1)p(V_2|V_1)$. This is very important for us, as $p$ consequently only expresses associative knowledge, but is ambiguous with respect to causality, i.e. the direction of influence.

In our treatment of IM, we look at possible ways that an agent can interact with their environment, which makes a causal perspective strictly necessary. In Appx. D, we do not learn BNs from observational data, but conversely, define a network to describe the underlying joint distribution. However, just as the joint distribution only conveys associational knowledge, so does the topology of a classic BN, thus allowing for distributions with potentially non-causal, spurious dependencies. Pearl (ibid., p. 21) observes that, when engineering BNs, AI researchers implicitly rely on knowledge of the causal processes underlying the involved variables to formulate reliable conditional independence assumptions and to keep the network sparse. He consequently identifies such assumptions as mere by-products of causal relationships, motivating the development of causal BNs which express causality directly in the network topology and in a specialised, interventional probability calculus (ibid., p. 22). A causal BN $C = (D, P_C)$ looks like a standard BN, but arrows $V_i \rightarrow V_j$ indicate direct causal effects from $V_i$ to $V_j$. Each conditional probability distribution in $P_C$ models a causal, stochastic mechanism which determines the value of $V_j$ based on the values of its parent variables.

This causal interpretation of probabilities relies on the concept of intervention, where we consider a conditioning variable to be actively set to a fixed value, rather than just being observed. For a query of the form $p(V_j|do(V_j = v_j), V_k = v_k)$, we thus want to find the distribution over values in $V_j$, given that we actively set $V_j = v_j$, expressed by the do operation, and observe $V_k = v_k$. The difference between observation and intervention is that the latter is done by an ‘external force’ and produces a mutilated network, in which the parents of the intervened variable $V_j$ are removed in the

---

1 This explains why in the literature, (classic) BNs are often treated synonymously to causal BNs, despite substantial differences (e.g. Bishop, 2006, p. 366).
Causal independence

Based on the calculus of intervention, we can complement stochastic and conditional independence conditions with their causal equivalents. For brevity, we indicate the variable to be intervened from now on with a dot, i.e. $\dot{X}$.

Ay and Polani (2008) write that $X$ and $Y$ are causally independent, imposing a third variable $Z$, if the following condition applies:

\[
(Y \perp X) \mid Z \iff p(y|x, z) = p(y|z) \quad \forall x, y, z \in \mathcal{X}, \mathcal{Y}, Z
\]

(B.5)

The interpretation of this is that intervening in $X$ after having intervened in $Z$ does not change the probability of observing $Y$. If we condition on the empty set, i.e. $Z = \emptyset$, we get the condition for (unimposed) causal independence:

\[
(Y \perp X) \iff p(y|x) = p(y) \quad \forall x, y \in \mathcal{X}, \mathcal{Y}
\]

(B.6)
If this condition applies, intervening in $X$ has no causal effect on $Y$. Note that in contrast to their standard probabilistic counterparts, these interventional conditions are, due to the directional nature of causality, not symmetric.

In Appx. C, we draw on the probabilistic foundations summarised in Appx. A to formulate the essential information-theoretical quantities used throughout this thesis. The notion of causal independence as introduced here plays a crucial role in the definition of information flow, e.g. to quantify the interaction of an agent with their environment.
Many intrinsic motivation (IM) models covered in this thesis, in particular empowerment maximisation (EM), are based on information theory. In this appendix, we introduce the underlying mathematical concepts. Information theory was originally introduced by Shannon (1948) to quantify and optimise the amount of information that can be passed through a noisy communication channel, e.g. in radio transmission. It is based on and extends probability theory in that it provides us with richer means to characterise and compare distributions and to examine probabilistic models. Our overview draws on standard literature by MacKay (2003) and Cover and Thomas (2006).

Classic information theory is free of semantics, both with respect to what establishes information in the first place, and to the label we assign to the agents involved in its communication. For instance, we might be interested in the message passing between two players in a game. Information theory allows us to quantify the information passed irrespective of the meaning of e.g. the individual words forming the messages. It is moreover agnostic with respect to the semantics that we might assign to the communicating agents: as engineers, we might consider them ‘players’ that exchange messages, but this label does not explicitly affect information-theoretic measurements on this system. We can thus apply the same concepts to other domains, e.g. to examine the information in DNA as it is reproduced from parent to daughter cells, subject to mutation (MacKay, 2003, p. 3). Its freedom of semantics allows for information theory to be applied across many disciplines, e.g. physics, economics, biology and machine learning.

In order to gain a good intuition for the information-theoretic quantities used in this thesis, we relate to their original use in communication. Fig. C.1 illustrates an abstract communication system, consisting of three main components: an encoder (transmitter), a decoder (receiver) and a communication channel. The ‘fundamental problem of communication’ (Shannon, 1948, p. 379) is to encode a message with sufficient systematic redundancy for it to be sent through a noisy channel and be recovered without ambiguity. This noise can be ambient, stem from physical properties of the channel, or it can be induced by other sources of uncertainty as introduced in Appx. A.

The encoder and decoder are modelled with conditional probability distributions, mapping messages $m$ to possible channel inputs $X$, and mapping channel outputs $Y$ to an estimate $\hat{m}$ of the original message. The fundamental problem of communication is approached by optimising the encoder and decoder mappings for efficient use of the channel. We focus on the special case of a discrete, memoryless channel. It is defined in terms of an input alphabet $X$, an output alphabet $Y$, and a set of conditional probability distributions $p(y|x)$, one for each element of the input alphabet $x \in X$. These distributions

---

1 However, a particular label might imply a different modelling of the underlying agent e.g. in terms of the encoding or decoding scheme $p(x|m)$ and $p(\hat{m}|y)$ used in communication. Information theory is agnostic with respect to the label, but sensitive to the agent modelling.
express the probability of observing a specific output $y$ given input $x$. The channel is *memoryless* because $p(y|x)$ only depends on the current input, and not on previous inputs or outputs. The measures introduced in the rest of this appendix can be used to quantify the information that can be expressed with a specific input- and output alphabet, and to assess how much information can be passed through the channel with an optimal or suboptimal encoding. In a more abstract sense, the measures are means to analyse the distribution of- and relationship between arbitrary random variables.

Information theory defines information as the reduction of uncertainty in the value of a random variable. The intuition behind this is that learning about a very unlikely outcome is more informative than learning about a likely one. This is captured in the concept of *self-information*:

$$I(x) = - \log p(x) \quad (C.1)$$

We use the binary logarithm by default and thus measure information in *bits*. The self-information satisfies three properties: (i) more likely outcomes convey less information and a guaranteed outcome provides us with zero information; (ii) less likely outcomes yield higher information; (iii) when $X \perp Y$, information is additive because $\log p(x, y) = \log p(x)p(y) = \log p(x) + \log p(y)$.

The self-information only considers a single outcome. The *entropy*, arguably the central measure in information theory, quantifies the uncertainty of an entire probability distribution as the expected amount of information in values drawn from that distribution:

$$H(X) = \mathbb{E}_{X \sim p}[I(x)] = \mathbb{E}_{X \sim p}[-\log p(x)] \quad (C.2)$$

$$= - \sum_x p(x) \log p(x) \quad (C.3)$$

By convention we assume that $0 \log 0 = 0$. The entropy represents a lower bound on the average amount of bits required to accurately describe the state of a random variable $X$ with the best possible compression. It is fundamental to communication theory, in that it gives us the shortest average length of a lossless message encoding based on the alphabet $\mathcal{X}$. In other words, it represents the limit to lossless compression. The entropy is non-negative and reaches its minimum at zero if the random variable is guaranteed to yield a specific outcome. It is maximum if $X$ is uniformly distributed, i.e. $X \sim \mathcal{U}$, and then simplifies to $\log |\mathcal{X}|$. Fig. C.2 illustrates this property for a binary variable, e.g. a coin flip, with $\mathcal{X} = \{0, 1\}$ and $p(X = 0) = 1 - p(X = 1)$.

The *joint entropy* of a set of random variables $X_1, X_2, \ldots, X_n$ follows directly by treating the set as a single variable in Eq. C.3:

$$H(X_1, X_2, \ldots, X_n) = - \sum_{x_1, \ldots, x_n} p(x_1, \ldots, x_n) \log p(x_1, x_2, \ldots, x_n) \quad (C.4)$$
Note that the entropy is free of semantics in that it only depends on the distribution of random variables and not on their specific values.

The conditional entropy measures the average amount of information required to describe \( Y \) if the value of \( X \) is known. It is given by the expectation over the entropies of conditional distributions:

\[
H(Y|X) = \mathbb{E}_{X \sim p}[H(Y|X)] = \sum_x p(x) H(Y|x) \\
= -\sum_x p(x) \sum_y p(y|x) \log p(y|x) \\
= -\sum_{x,y} p(x,y) \log p(y|x) \\
= -\mathbb{E}_{X,Y \sim p}[\log p(y|x)]
\]

(C.5)

(C.6)

(C.7)

(C.8)

Here we applied Eq. A.15 to Eq. C.6 to rewrite the conditional entropy as an expectation over the joint distribution of \( X \) and \( Y \). Note that \( H(Y|X) = 0 \) if the value of \( Y \) is completely determined by \( X \). Furthermore, \( H(Y|X) = H(Y) \) if and only if \( Y \) and \( X \) are independent random variables, i.e. \( X \perp \!\!\!\!\!\perp Y \).

The chain rule for entropy allows us to write the joint entropy as the sum of conditional entropies:

\[
H(X_1, X_2, \ldots, X_n) = \sum_{i=1}^n H(X_i|X_{i-1}, X_{i-2}, \ldots, X_1)
\]

(C.9)

Applied to a pair of random variables, we get the two alternative expressions:

\[
H(X,Y) = H(Y) + H(X|Y) \\
= H(X) + H(Y|X)
\]

(C.10)

(C.11)

The entropy of two variables thus corresponds to the entropy of one variable plus the conditional entropy of the other. If \( X \perp \!\!\!\!\!\perp Y \), we find that \( H(X,Y) = H(X) + H(Y) \) because \( H(Y|X) = H(Y) \) as shown above.
The previous measures apply to a single distribution over one or a set of random variables. The relative entropy (or Kullback-Leibler divergence) in contrast measures the pseudo-distance between two distributions $p(x)$ and $q(x)$ defined on the same state space $X$:

$$D(p||q) = E_{X \sim p} \left[ \log \frac{p(x)}{q(x)} \right] = \sum_x p(x) \log \frac{p(x)}{q(x)}$$

(C.12)

For this definition, we use the convention $0 \log \frac{0}{0} = 0$ and $p \log \frac{p}{0} = \infty$. We can interpret the relative entropy as the inefficiency of assuming $p$ if the actual distribution is $q$. It is non-negative, and zero if $p = q$. In communication, it represents the expected number of extra bits required to encode samples from $p(X)$ using a code optimised for $q(X)$. Importantly, the relative entropy is not a proper distance metric because it is not symmetric, i.e. $D(p||q) \neq D(q||p)$.

The mutual information measures the amount of information shared between two random variables $X$ and $Y$. It is the relative entropy between the joint distribution $p(X, Y)$ and the factorisation $p(X)p(Y)$:

$$I(X; Y) = \sum_{x,y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)}$$

(C.13)

$$= \sum_{x,y} p(x, y) \log \frac{p(x|y)}{p(x)}$$

(C.14)

Here, we use the convention $0 \log \frac{0}{0} = 0$ and $p(x, y) \log \frac{p(x|y)}{0} = \infty$, and Eq. C.14 results from applying Eq. A.9 to Eq. C.13. By measuring the distance between the left- and right-hand side of the two equivalent stochastic independence conditions in Eqs. A.10 and A.12, the mutual information quantifies how much the variables $X$ and $Y$ deviate from stochastic independence. We can rewrite the mutual information as a difference of entropies:

$$I(X; Y) = \sum_{x,y} p(x, y) \log \frac{p(x|y)}{p(x)}$$

(C.15)

$$= \sum_{x,y} p(x, y) \log p(x|y) - \sum_{x,y} p(x, y) \log p(x)$$

(C.16)

$$= - \sum_x p(x) \log p(x) - \left( - \sum_{x,y} p(x, y) \log p(x|y) \right)$$

(C.17)

$$= H(X) - H(X|Y) = H(Y) - H(Y|X)$$

(C.18)

Here, we have applied basic logarithmic identities and marginalised the right-hand joint distribution in Eq. C.16 to get the entropy in Eq. C.17. The mutual information is symmetric, i.e. $I(X; Y) = I(Y; X)$, and we get $I(X; X) = H(X)$. Moreover, it is also non-negative because $H(X|Y) \leq H(X)$. Fig. C.3 illustrates the relationship between different entropies and the mutual information.

The mutual information is zero if $X$ and $Y$ are stochastically independent, and larger otherwise. We can furthermore quantify the deviation from con-
ditional independence, given a third variable $Z$, by means of the conditional mutual information which marginalises over the conditioning variable:

\[
I(X; Y|Z) = \sum_z p(z) \sum_{x,y} p(x, y|z) \log \frac{p(y|x, z)}{p(y|z)}
\]

\[
= \mathbb{E}_Z I(X; Y|z)
\]

Applied to two (sets of) variables in a classic BN, the mutual information measures their non-linear correlation, but is not sensitive to the direction of influence, i.e. to causality between the variables.

As a crucial ingredient for the definition of empowerment maximisation (EM), the central model of IM in this thesis, we need a means to measure how information is processed in a system. In other words, we require a means to quantify the (directional) information flow between random variables $X, Y$ as the amount of information about $X$ that is causally transmitted from $X$ to $Y$. To capture this directionality, we rely on causal BNs and interventional probability distributions. In analogy to the conditional mutual information in Eqs. C.19 and C.20, Ay and Polani (2008) quantify information flow by measuring the relative entropy between the left- and right-hand side of the causal independence condition in Eq. B.5. They define the information flow in a causal BN between variables $X$ and $Y$, imposing a third variable $Z$, as:

\[
I(X \to Y|Z) = \sum_z p(z) \sum_{x,y} p(x|z) \sum_y p(y|x, z) \log \frac{p(y|x, z)}{\sum_{x'} p(y|x', z) p(x'|z)}
\]

\[
= \sum_z p(z) \sum_{x,y} p(x, y|z) \log \frac{p(y|x, z)}{p(y|z)}
\]

\[
= \mathbb{E}_Z I(X \to Y|z)
\]

The differences between this and Eqs. C.19 and C.20 are rather subtle. The standard mutual information is based on probability distributions condi-

---

\footnote{Information flow is often confused with the transfer entropy, but the quantities only coincide in specific situations. Cf. Lizier and Prokopenko’s (2010) analysis for a thorough distinction.}

---

Figure C.3: Venn diagram describing the relationship between different entropies and mutual information. Adopted from Cover and Thomas (2006, p. 20).
tioned on observations. The information flow in contrast quantifies the distance between the two distributions resulting from the interventional queries \( p(y|x, z) \) and \( p(y|z) \); it becomes zero if intervening on \( Z \) makes no difference to the distribution of \( Y \), thus confirming causal conditional independence \( (Y \perp X) \mid Z \) (cf. Eq. B.5). If we condition on the empty set, i.e. for \( Z = \emptyset \), we get \( I(X \to Y) \), the deviation from (unimposed) causal independence \( (Y \perp X) \) (cf. Eq. B.6). If \( Z = V \setminus (X \cup Y) \), with \( V \) being the set of all nodes in the network, we quantify the direct (unmediated) causal effect of \( X \) on \( Y \); it is only larger than zero if there is a causal effect between \( X \) and \( Y \) which is not mediated by any other set of variables in \( V \). In contrast to the mutual information, information flow is not symmetric.

A communication channel (Fig. C.1) represents a causal mechanism between input \( X \) and output \( Y \). We thus express this potentially noisy mapping with the interventional distribution \( p(Y|X) \). The information flow then corresponds to the average amount of information transmitted through the channel. By applying Eq. C.18, we can understand the information transmitted through a given channel as the average uncertainty in the input \( X \), reduced by the knowledge of a specific output \( Y \). We thus measure the amount of information which the received signal on average contains about the transmitted one. This reflects the prior definition of information as the reduction of uncertainty.

Channel Capacity

This measure for information flow quantifies how much information on average passes through a given channel for a fixed input distribution \( p(X) \). The channel capacity for a memoryless channel corresponds to the maximum amount of information that we can transmit for any specific distribution, chosen from all possible input distributions:

\[
C = \max_{p(x)} I(X \to Y) \tag{C.24}
\]

The channel capacity measures the ‘maximum amount of error-free information that can be transmitted over the channel per unit time’ (MacKay, 2003, p. 149), thus answering another fundamental question of communication theory. Calculating the channel capacity requires to find the optimal \( p^*(X) \) (we indicate optimal solutions with an asterisk) which induces the maximum information flow for the given interventional channel distribution \( p(Y|X) \). Given such a distribution, the iterative Blahut-Arimoto algorithm (Arimoto, 1972; Blahut, 1972) yields the channel capacity for a memoryless channel with arbitrary precision, albeit at a high computational expense.

In the final Appx. D, we apply the mathematical concepts introduced in Appxs. A and B to formalise the interaction of an agent with their environment in a way which allows for information-theoretic analysis, and thus for the formalisation of various models of IM, including empowerment maximisation (EM) as the central model in this thesis.
Here, we introduce the perception-action (PA)-loop as a specific model of an agent’s interaction with their environment, which is used throughout this thesis, in corresponding publications, and in related work. Our formalisation draws on joint work (Biehl et al., 2018), but both the concept and its original formalisation date further back. We begin by sketching this heritage.

When formalising an agent-environment interaction as a PA-loop, one (i) considers neither the agent nor the environment as passive, but as coupled systems that causally influence each other in a temporally expanded, closed-loop interaction: an agent performs actions which contribute to changes in their environment, which in turn are (partially) perceived through their sensors, leading to changes in agent-internal states, e.g. memory, and causing new actions. Vice versa, the environment impacts on the agent’s sensor, causing a reaction which contributes to changes in the environment. This view (ii) expresses a holistic understanding of the environment dynamics and the behaviour of the agent; neither can be understood in an isolated way. At the same time, it distinguishes agent and environment, thus (iii) emphasising an agent’s embodiment as a unique and limited coupling with the world, (iv) while allowing for a loose specification of the agent’s boundary, a task which remains problematic (cf. Biehl, Ikegami & Polani, 2016).

The ideas inherent to the concept of a PA-loop can be traced back to at least the 19th and early 20th century. The physiologist Sechenov proposes a holistic perspective on an agent and their environment as early as 1861:

> ‘The organism cannot exist without its supporting external environment; hence a scientific definition of the organisms should include also the environment which influences it.’
> (Sechenov, 1965, p. 122; as cited in Lagerspetz, 2001)

The biologist and early cyberneticist von Uexküll strengthens this view by opposing the mechanistic conception of biology of his time, which understood organisms as passive subjects that conform to the laws of material causality. Influenced by Kant (1790/1995, § 64), he considers organisms active agents:

> ‘But we who still hold that our sense organs serve our perceptions, and our motor organs our actions, see in animals as well not only the mechanical structure, but also the operator, who is built into their organs as we are into our bodies. We no longer regard animals as mere machines, but as subjects whose essential activity consists of perceiving and acting.’ (Von Uexküll, 1992, p. 6).

Von Uexküll (1920) has developed a non-anthropocentric theory of meaning, allowing us to understand organisms as embodied agents that respond and act in their world in ways that are meaningful from their own perspective. Based on his empirical studies of reflex arcs in marine invertebrates, he has
shown that organisms engage in circular, functional interactions\(^1\) with their environment (Von Uexküll, 1920, pp. 97), thus shaping the understanding of the agent-environment interaction as a coupled, closed-loop system. This holistic perspective and his agent-centric account of sense-making have influenced cybernetics (e.g. via Ashby, 1954, as discussed by Lagerspetz, 2001), ecological theories of perception (e.g. Gibson, 1979), and embodied (Brooks, 1991) as well as theories of enactive cognition (Maturana & Varela, 1991).

These theories in turn have influenced the formalisation of the PA-loop e.g. by Beer (1995), who has modelled the closed-loop interaction of an embodied agent with their environment as two coupled, dynamical systems. Agent and environment are considered mutual sources of perturbation which affect their joint future interaction. Of particular interest for us is Touchette and Lloyd’s (2004) formalisation of closed-loop control as a (causal) BN. With the goal to determine information-theoretic limits to optimal control and observation, they have interpreted the arrows between sensor, actuator and a controlled variable as communication channels\(^2\), which allows for the analysis of information flow between these variables. However, they only model 1-step control, without reference to the concepts of agent or environment\(^3\). Klyubin, Polani and Nehaniv (2004) have reconciled Beer’s dynamic systems model with Touchette and Lloyd’s information-theoretic framework by assigning their abstract control components the semantics of an agent interacting with their environment, and by unrolling this interaction in time. In addition, they have introduced an agent’s memory, which can also serve as a placeholder for other agent-internal components. Their formalisation has been used in subsequent studies (e.g. Klyubin, Polani & Nehaniv, 2007; Bertschinger et al., 2008; Klyubin, Polani & Nehaniv, 2008; Ay et al., 2012; Ghazi-Zahedi & Rauh, 2015; Biehl & Polani, 2017), and our account is derived from it.

---

\(^1\) Some interpret these functional cycles ('Funktionszyklen') as mere sensorimotor loops, while others consider them inseparable from Uexküll’s theory of meaning (cf. De Jesus, 2016). This distinction does not matter to us, as functional cycles would subsume sensorimotor loops.

\(^2\) The analogy between control and communication has been pointed out by other authors before, but it has only been formalised by Touchette and Lloyd in 2000. The information-theoretic treatment of sensors and actuators is already implicit in Ashby’s (1956) earlier work.

\(^3\) In contrast to Beer (1995), Touchette and Lloyd’s (2000) control-theoretic perspective does not relate to the semantics of an agent interacting with their environment. In particular their ‘controlled system’ could be a part of a larger environment which is not accounted for as a whole. The PA-loop as formalised here accounts for the whole environment.
We first formalise the PA-loop for the slightly more complex case where the agent has memory, and then simplify it. We assume that the agent is part of a larger agent-environment system which is discrete in space (i.e. its variables take on discrete values) and discrete in time. We then define the PA-loop as a causal BN $C = (V_C, P_C)$, with the graph structure shown in Fig. D.1a.

We assume that the agent-environment system originates at some ‘big bang’ moment $t = 0$. The figure shows a brief slice of two time-steps from the initialisation onwards, and a further two-step slice later on with generic time indices. The set of nodes $V_C$ represents a partitioning of the overall agent-environment system into the following random variables:

- Agent sensor $S$ with state space $\mathcal{S}$
- Agent memory $M$ with state space $\mathcal{M}$
- Agent actuator $A$ with state space $\mathcal{A}$
- The rest of the system $R$ with state space $\mathcal{R}$

This partitioning realises an agent-centric perspective where everything that is not captured by the agent’s sensor, memory and actuator is encapsulated in the rest of the system or agent-external environment. By ‘sensor’ and ‘actuator’, we mean the values returned by or fed into either a physical or virtual sensor or actuator, respectively. Recall from Appx. B that the nodes of a BN can represent single, but also vector-valued and sets of random variables. Thus, $S$ can encompass several different sensors, $A$ different actuators, etc.

The set $P_C$ describes the causal dependencies between these random variables by means of interventional probability distributions (cf. Appx. A).

- Sensor dynamics $p(s_t|\mathcal{R}_t)$
- Memory dynamics $p(m_t|s_t, m_{t-1}, a_{t-1})$
- Initial memory dynamics $p(m_0|s_0)$
- Action policy $p(a_t|m_t)$
- Environment dynamics $p(\mathcal{R}_{t+1}|a_t, \mathcal{R}_t)$
- Initial environment state $p(\mathcal{R}_0)$

They describe the following causal dynamics: At time $t = 0$, the rest of the system is initiated from a state $\mathcal{R}_0$, which is then perceived through the agent’s sensor $S_0$. The latter influences the agent’s memory state $M_t$. At any subsequent time $t > 0$, the agent’s memory is additionally shaped by the past memory state $M_{t-1}$ and past action $A_{t-1}$. The agent chooses a next action $A_t$ based on its memory $M_t$. This action then affects how the rest of the system transitions from its current state $\mathcal{R}_t$ to its new state $\mathcal{R}_{t+1}$. The new state again impacts on the sensor, and the system keeps looping indefinitely. These distributions realise the Markov property, i.e. the state of the system at time $t + 1$ only depends on the system’s state at the prior time-step $t$.

To fully define the PA-loop, we have to specify all state spaces and interventional distributions given above, and we do so for each individual experiment.
in the corresponding chapter. We assume that state spaces and interventional distributions are time-homogeneous, i.e. that they are invariant with respect to the current time-step. The dynamics at \( t = 0 \) are excluded from this assumption. A common, additional assumption is that the initial state \( r_0 \) of the rest of the system is given. Its distribution can then be specified by Kronecker’s delta (cf. Eq. A.7), i.e. \( p(r_i) = \delta_{r_i,r_0} \forall r_i \in \mathcal{R} \). Given time-homogeneity and a full specification of the PA-loop, we can use Eq. B.3 to define the joint distribution for the time slice \( 0, \ldots, T \). To this end, we use the shorthand \( r_{\leq T} \) for the sequence of states \((r_0, r_1, \ldots, r_t, r_{t+1}, \ldots, r_T)\):

\[
p(r_{\leq T}, s_{\leq T}, a_{\leq T}, m_{\leq T}) = \prod_{t=1}^{T} p(a_t|m_t)p(m_t|s_t, m_{t-1}, a_{t-1})p(s_t|r_t)
\]

\[
\times p(r_t|r_{t-1}, a_{t-1})
\]

We use the previous formalism for a generic formalisation of EM as our central model of IM. For our applied contributions though, we make the simplifying assumption that agents do not have memory, with the resulting PA-loop shown in Fig. D.1b. The causal dependencies in \( P_C \) are:

- Sensor dynamics \( p(s_t|r_t) \)
- Action policy \( p(a_t|s_t) \)
- Environment dynamics \( p(r_{t+1}|a_t, r_t) \)
- Initial environment state \( p(r_0) \)

With these simplified causal dependencies, the factorisation of the joint distribution also simplifies to:

\[
p(r_{\leq T}, s_{\leq T}, a_{\leq T}) = \prod_{t=1}^{T} p(a_t|s_t)p(s_t|r_t)p(r_t|s_{t-1}, a_{t-1})
\]

\[
\times p(a_0|s_0)p(s_0|r_0)p(r_0)
\]

The formalisation of the PA-loop as a causal BN (Klyubin, Polani & Nehaniv, 2004) comes with a range of benefits. Its causal probabilistic structure allows for the consistent measurement of information-theoretic quantities, specifically information flow (cf. Appx. C): we can trace how information is captured by the agent’s sensors, persisted in memory, and fed back into the environment via their actions, to be captured again later (Klyubin, Polani & Nehaniv, 2007). We can thus analyse the effect of sensing on acting and vice-versa in a unified framework. The PA-loop complements the universality of information theory (cf. Appx. C) in that the abstraction into constraints on the agent’s embodiment is agnostic with respect to a specific information processing architecture, e.g. receptive fields or layers of neurons (ibid.). The PA-loop can be used to model the true ‘physical’ dynamics underlying the agent-environment system, but it can also serve as an epistemic, subjective model of these dynamics, estimated by the agent.
The PA-loop provides us with two perspectives on an agent’s embodiment. Ziemke (2003) points out substantial ambiguity in how this term is understood, and we adopt the arguably broadest view that a system is embodied if it is ‘structurally coupled’ to their environment (Maturana & Varela, 1987). We extend this to what Ziemke (2003) refers to as sensorimotor embodiment, where the coupling is facilitated by the sensors and actuators of either a virtual or physical agent body. This coupling is formally given by:

\[ p(s_{t+1}, \ldots, s_T|a_t, r_t) \]  

(D.3)

Note that we do not account for any other ‘perturbatory channels’ (Quick et al., 1999, p. 2) between the agent and the environment other than their sensors and actuators (cf. Ghazi-Zahedi & Rauh, 2015, for a more comprehensive account). To fully specify the embodiment, we need to define the state spaces \( S, A \) and \( R \) as well as the sensor and environment dynamics. Since the state of the world \( R \) is not directly accessible to an agent, this corresponds to an objective view of their embodiment. In our definition of EM in Sec. 3.2, we distinguish a subjective perspective for which an agent constructs their embodiment by inference. (Eqs. 3.5–3.7).

We highlight further properties of the PA-loop by comparing it against partially observable Markov decision processes (Sutton & Barto, 2018, p. 466), a formalism that especially reinforcement learning (RL) researchers might be more familiar with. The definition of such processes does not entail memory dynamics, and crucially, no action policy. They describe a sequential decision problem to be solved by an optimal policy that maximises reward. This solution is not part of the process definition. The PA-loop, more holistically, defines the concrete interaction between an agent and their environment, by incorporating their specific action policy and memory dynamics. As a minor difference, a partially observable Markov decision process defines an intrinsic or extrinsic reward function (cf. Barto, 2013; and Ch. 2) and a discount factor explicitly, both of which can be implicitly captured in the PA-loop’s policy definition.


Cook, M., Colton, S. & Gow, J. (2016b). The ANGELINA Videogame Design System – Part II. Trans. on Computational Intelligence and AI in Games, 9(3), 254–266.


Workshop at the Artificial Intelligence and Interactive Digital Entertainment Conference (AIIDE), 31–38.


Hello Games. (2016). *No Man’s Sky*.


