Sheldon, Kennon M.; Martela, Frank

A modest proposal

Published in:
JOURNAL OF POSITIVE PSYCHOLOGY

DOI:
10.1080/17439760.2021.2016901

Published: 04/03/2022

Document Version
Peer reviewed version

A Modest Proposal: Free Will is Real

Kennon M. Sheldon [1,2] and Frank Martela [3]

[1] University of Missouri

[2] Russian National University, Higher School of Economics

[3] Aalto University

This research was supported within the framework of a subsidy by the Russian Academic Excellence Project '5-100'

This is a post-print version of the manuscript published as:
Sheldon, K. M., & Martela, F. (2022). A modest proposal: Free will is real. *The Journal of Positive Psychology*, *17*(2), 271–280. https://doi.org/10.1080/17439760.2021.2016901

Abstract

We present a hybrid philosophical/psychological argument in defense of "free will" – namely, the human capacity to influence our own destinies, via our goal choices. The argument builds on the proposals in philosopher Christian List's (2019) book, *Free will is real*. We show that List's psychological account of free will -- that it requires the ability to consider alternatives, to make a choice, and to enact that choice -- has already been proven, via research into human prospection and goal functioning. Thus, the real question is: "can we use our free will *wisely*?" Self-concordance research and implicit/explicit motive discrepancy research suggest that sometimes, we can't. Ironically, the same fact that makes us free -- that our choices are based on our imaginative capacities -- also makes us free to be clueless, i.e., out of touch with ourselves or with reality. In building our argument for free will we also discuss self-determination theory, the evolved symbolic self, system 1 versus system 2 functioning, and the hierarchical organization of human existence.

Martin Seligman has made many important contributions to the study of active agency – to our understanding of how people can learn optimism, imagine new possibilities, and achieve remarkable things. We would like to express our appreciation for Dr. Seligman's life and work by pushing the concept of agency to its limits, asking: "do we human beings have free will, in any meaningful sense? Or instead, are our behaviors always determined by inscrutable factors that we can never know about?" In other words, what are the current psychological arguments for the existence of free will, and for our subjective feelings of being "the deciders" within our own lives, constrained and ineffective as we may (yet) be?

Obviously, free will is a hugely contentious topic, with vast literatures concerning both the psychology (e.g. Baumeister, 2008; Baumeister, Masicampo, & Vohs, 2011) and philosophy (e.g. Kane, 2011, List, 2019) of the question. We cannot even begin to do justice to these literatures, in this short article. Instead we will try to outline one way forward, already supported by research data. We'll show how this way avoids many of the perennial logjams and quagmires concerning free will, while also providing important prescriptions for positive psychology, and specifically, the question of how people can best live their lives. The arguments are mainly condensed from Sheldon's forthcoming (2022) book on the topic of free will.

Our approach begins with conditional acceptance of philosopher Christian List's proposals, in his (2019) book *Why free will is real* and related articles (List, 2014, 2019). Side-stepping much of the philosophical terrain, List adopted a *psychological* approach to free will, based on the concept of intentionality. By List's definition, free will requires three related capacities: 1) the capacity to generate and evaluate alternative possibilities for behavior; 2) the

capacity to make a choice among these alternatives; and c) the capacity to take action to bring about the chosen alternative. Thus, free will merely requires an intelligent control system (potentially including an artificial one), which can generate alternatives, then decide what it wants, and then pursue that decision. Seligman, Railton, Baumeister, and Sripada (2013, 2016) took a similar position on the free will question, by focusing on the human capacity for *prospection* -- the ability to imaginatively foresee the results of various behavioral alternatives, in order to better decide among them. Seligman and colleagues argued that prospection is one of the key psychological capacities required for the exercise of free will. We agree, and attempt to integrate prospection into a broader argument for the existence of free will.

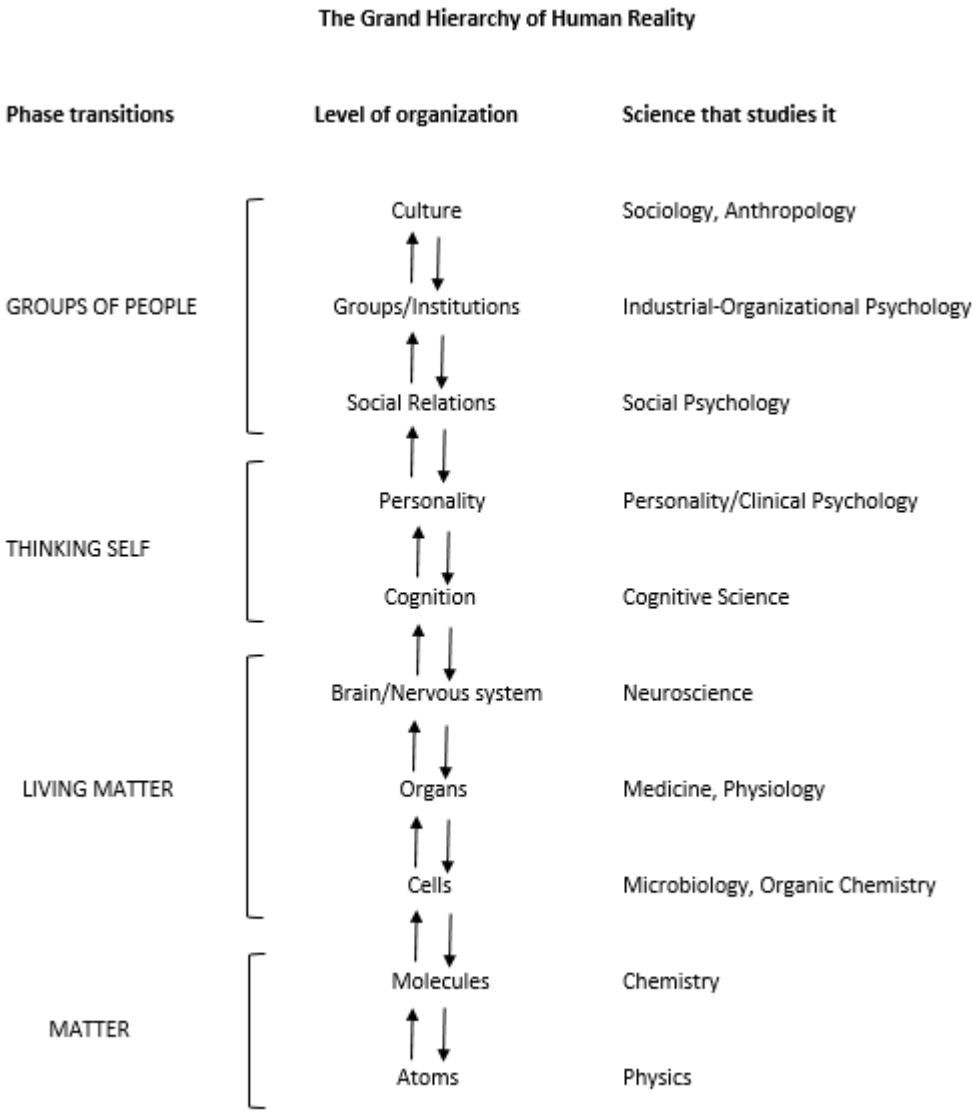
How free will can exist in a deterministic universe

But can human control systems really "decide what they want," independently of the physical substances and mechanisms that constitute them? In important ways, yes. Our basic proposal, much indebted to List (2014, 2019), is that there can be deterministic causality lower down within a hierarchically organized system, as at the level of cells, organs, and neurotransmitters, while at the same time, there can be true agentic choice and non-deterministic causality higher up in that system, as at the level of minds and networks of minds (Sheldon, 2004). This doesn't require uncaused causes, magic, the existence of souls, or anything else non-scientific: it just requires recognition that there are many irreducible levels of organization and explanation in the universe, including personhood (Ryan & Deci, in press).

Thermodynamical laws, for example, reveal patterns in the behavior of liquids and gases that are not visible at the atomistic level of explanation. While the higher level patterns of organization are dependent and built upon the lower levels, they can also exhibit dynamics that

are not simply reducible to lower-level mechanisms. We will argue that self-organizing human minds, cast as symbolic selves partially cut off from their own implicit processes, and operating within highly complex self-narratives and information environments, have very large degrees of freedom to determine their own behavior.

Thus, in grappling with the question of whether (or how) free will exists, we believe it is crucial to understand the distinction between the many levels of organization and explanation, as is increasingly recognized in contemporary philosophy of science and metaphysical theories (Kim, 2002; List, 2019; Oppenheim & Putnam, 1958; Owens, 1989). In their classic treatment of the topic, Oppenheim and Putnam (1958) proposed six levels: Elementary particles, atoms, molecules, cells, multicellular living organisms, and social groups, with each level having its own entities, laws, and dynamics. Sheldon (2004, 2011) posited a similar hierarchy which fills in more layers, as shown in the diagram below. The diagram illustrates several important phase transitions: from mere matter to living matter, from living matter to mind, from mind to self, and from one self in one body, to networks of selves in many bodies.



Upward causation and emergence

Starting at the bottom of the diagram, the upward-pointing arrows show that we are each constituted (ultimately) by atomic particles, which in turn are nested inside of (constitutive of) molecules, which are nested inside of cells, inside of organ systems, inside of nervous systems, inside of cognitive processes, inside of personality processes, inside of relational, group, and cultural processes. The lower levels of organization support the levels

above, such that higher levels could not exist without them. This is why malfunctions at lower levels can disrupt higher-level functioning – for example, damage to the brain tends to disrupt psychological functioning. Activity can occur at all levels simultaneously, although the time scale for unfolding events is usually slower, higher up in the hierarchy. At each level, things are happening which are best understood *at* that level; the effects can rarely be completely reduced to or explained through the lower-level processes. This is because each new level of organization is more complex than the ones below, in that it takes the lower levels as a set of building blocks, upon which it adds its own unique potentials and patterns.

Again, it is nowadays quite widely accepted by scientists and philosophers alike that higher levels of organization can exhibit "emergent" properties that are "*not explainable* in terms of – or, alternatively, *not reducible* to – lower-level properties" (List, 2019, p. 870). The higher-level phenomena *supervene* on the lower-level phenomena in the sense of being constituted by the lower-level phenomena, requiring them for their own existence, but also adding something more. And, "If we were to dispense with higher-level descriptions, we would fail to see some higher-level patterns in the world." (List, 2019, p. 873.) Thus, instead of endorsing a naïve flat ontology which holds that the world can be exhaustively described at one 'fundamental' level, scientists need to recognize that different levels require irreducibly different concepts, entities, and theories to provide predictive models applicable at that level.

Indeed, this is why new sciences were continually invented during the last 400 years, each with a new vocabulary and a new way of thinking about causation. From the bottom up (as shown in the diagram), first came physics, then chemistry, then organic chemistry, then cellular biology, then physiology, then neurology, then psychology, then sociology and

anthropology. At the upper levels of the hierarchy the action is increasingly "mind-like," that is, it is increasingly focused on the manipulation of information rather than matter. Of course, this information does not exist independently of matter – it requires matter, and the biological operations that take place in physical brains.

Nevertheless, accepting that the symbolic operations of the mind have causal power to affect human behavior -- thus amounting to a *top-down causation* from mind to matter -- seems to lead to many useful predictions about human behavior. Such top-down causation might not seem to be 'real' from *the point of view* of that lower level: Each individual molecule still adheres to the laws of physics, with no break in the physical chain of causation. But *from the point of view* of the higher level, a psychological theory (such as the fight-or-flight response; Fuller et al., 2010) might be predictive of what the corpus of cells in an organism does next (such as running away in a particular case).

As an analogue, think about water inside a glass flask starting to boil, and the pressure created by this boiling breaking the glass. This, according to List, is a situation where the most meaningful causal explanation takes place at a higher thermodynamical level, rather than at the atomic level. While each individual molecule taken individually will behave according to the natural laws applicable at that level, their aggregate behavior – the water molecules now being puddled on the floor, instead of being contained inside the bottle – is most meaningfully causally explained by the higher-level thermodynamical laws. Similarly, the most meaningful causal explanations of how the aggregation of particles that is a human body will behave is not provided at the atomic level, but at the level of the intentions of an agent with free will.

Top-down control and downward causation

Turning to the *downward*-pointing arrows in the figure: these represent the fact that from the point of view of the higher levels, there are top-down causal effects on the systems that support them. Things are happening up there that affect the destiny of the systems below. As one example, the governing process of a corporation (near the top of the hierarchy) can reach a decision that impacts thousands of workers contained down within that corporation – and of course, all the cognitive and physical processes contained within the body of each worker. Similarly, a person can make a decision that impacts the subsequent functioning of the trillions of cells contained within the body of that person. If we want to predict where that person's cells will be located five minutes from now, physical theories will not help much. But if we hear our friend say over the phone that 'I will be there in five minutes', then we can predict that those cells will be moving towards us soon.

Free will is thus not a concept applicable at the lower levels. Looking at the levels of atoms and molecules, it is hard to see what 'free will' could amount to there. Molecules don't have a mind or will that could or could not be free. When we talk about free will, we are necessarily talking about a higher-level concept, operational mainly at the psychological level. As List (2014, p. 156) notes, "free will is not a physical phenomenon, but a higher-level one on a par with other higher-level phenomena" – such as beliefs, desires, and feelings. Free will is about *agential possibility*, the ability of the agent to do otherwise, and thus to make choices about what course of action to take. Accordingly, the "appropriate frame of reference", when talking about free will, "is not the one given by fundamental physics, but rather the one given by our best theories of human agency" (List, 2014, p. 161). In talking about free will, we are thus strictly speaking talking about *agential free will,* the ability of the agent to do otherwise

and have a sense of choice, which are concepts applicable only at the behavioral and psychological level. And indeed, the concept of an agent or executive function capable of making choices is at the heart of psychological, economic, decisional, sociological, and anthropological theories of human behavior. In aiming to understand and generate predictive models of human behavior, within models containing large degrees of freedom – which is what behavioral sciences in general do – it is virtually impossible to circumvent the assumption of human agency. So, as long as we operate on the higher level of explanation, free will (peoples' ability to choose among alternatives) is one of the fundamental proposals we need to make.

Control systems: How human intentions gain causal force

Often, regnant processes at a particular level operate in the form of *control systems*, in which set-points and goal-standards are used to regulate more basic processes occurring lower down in the hierarchy. Control systems can be found at every level of life. At the bottom layer, cell membranes regulate the transport of molecules into and out of the cell. Higher up, the kidney regulates fluid and mineral balances within the blood. At the level above that, the nervous control system regulates the kidney's set-point functioning. Far higher up, at the level of society, institutional control systems regulate conformity to group laws, standards and norms.

Indeed, the evolution of life, in a certain sense, has always involved the introduction of increasingly complex control systems. Simple bacteria acquired the ability to react to the pH levels of their environments, in the sense that they have regulatory mechanisms leading them to move towards environments with more optimal pH levels from the point of view of their functioning. Animals evolved even more advanced control systems to regulate their behavior,

such as the homeostatic systems leading them to seek water when dehydrated and food when hungry. What is unique about humans is that we have developed a capacity to think in the long term (Roberts, 2002, and to deliberately choose behavioral targets far in the future, and then put in the necessary work to achieve them. A human being is a "prospecting organism", and this capability to act based on evaluations of future possibilities has a crucial evolutionary advantage for our species (Seligman et al., 2013 p. 120). Instead of a mere concern with short-term physical needs, humans can set abstract long-term goals and regulate their behavior to achieve those goals.

We propose that, at the level of personality processes, the relevant control systems are the *mental intentions* that we have formed, which we use to guide and regulate the functioning of our minds and action systems (Carver & Scheier, 1998). We now know that human beings are constantly thinking about their "current concerns" in life (Klinger, 1977), are characteristically striving for many kinds of things (Emmons, 1989), and are continually thinking up new projects for themselves (Little, 2011). Personality research at the level of "characteristic adaptations" (McCrae & Costa, 1996; McAdams, 1996) has come a long way in showing how such goals and projects can literally take the reins of the brain, using top-down regulation and negative feedback processes to accomplish all kinds of things. Personal goals are how we organize (or try to organize) ourselves in the long-term, bringing about (or failing to bring about) the lives that we think we want.

Notice that the description of human behavior provided by personal goal research includes precisely the capacities that Christian List defines as constituting "free will:" that people have the capacity to imagine and consider behavioral alternatives, to make a selection

among those alternatives, and to take action based on their selection. Research on goal-functioning shows that people are constantly crossing "Rubicons" -- moving from inaction to action, prompted merely by ideas which have appeared within their minds, which they have decided to follow (Gollwitzer, 2012; Sheldon, Prentice, & Osin, 2019). Goal intentions work much the same way that control standards do lower down in the body, in that they set the targets that the system tries to enact, through negative and positive feedback. But there is much more flexibility in mental control systems, and much less automaticity, compared to physiological control systems; many more possible alternatives considered, and many more degrees of freedom. It is these very large degrees of freedom that provide humans with much of their potential free will, and gives them the ability to live startling and creative lives.

From *whether we have* free will to how *well* we *use* our free will

Still, there is a critical issue that is left unaddressed by List's account: the issue of *wise* choice. Suppose we accept that people have free will because they have intentionality, which lets them consider possibilities, then choose and enact particular intentions. But what if they make *bad* choices – selecting goals and intentions that won't help them, won't make them happy, and that might even be harmful to them? Or what if they are choosing based on *other* peoples' perceived demands, rather than on their own heartfelt desires or growth impulses? What if they feel coerced in their intentions, as if they have no real freedom to choose? Or what if they simply have no idea who they really are, or what will actually make them happy? Negative freedom, the condition of having no external constraints on my choice, is not enough for positive freedom, the sense that I am able to take control of my own life according to my self-chosen purpose and values (Berlin, 1969). Free will in the fullest sense requires not only

negative but also positive freedom. The question is: How to best exercise our positive freedom, to make apt choices from among well-specified alternatives?

Sheldon's "self-concordant goals" research (Sheldon & Elliot, 1999; Sheldon, 2014; Sheldon & Goffredi, in press) has shown that selecting apt personal goals can be quite difficult. Many people write down goals and intentions that do not seem to fit them, and that fail to bring them satisfaction and fulfillment even when they are achieved. It is as if they don't know how to use their free will capacity, or how to use it in a personally beneficial way. The self-concordance model is based on Self-determination theory (SDT; Deci & Ryan, 1985; Ryan & Deci, 2017), which says that not all intentional behaviors are *experienced* as autonomous and self-determined – sometimes we feel very UN-free, or controlled, in our intentional behaviors. Indeed, this is the main factor that distinguishes SDT from other theories of active agency – SDT's emphasis on the endorsement of the person's behavior by the phenomenal self. SDT shows that feeling autonomous is a major problem for many people, given that they must continually interact with powerful others who try to control them or to compel them to do things they don't want to do. In terms of the grand hierarchy, other people (and the group functions they represent) can try to exert downward causality upon us, introducing unwanted constraints upon our free will.

Self-concordance (conceptually defined as person-goal fit) is typically measured by asking people "why" they are pursuing the set of goals they have just written down, using the reasons specified by the organismic integration mini-theory of Self-determination theory (SDT; Ryan & Deci, 2017). The assumption is that goals chosen based on feelings of pressure and compulsion are probably a worse fit for the striver, compared to goals based on feelings of

enjoyment, engagement, and identification. This matching hypothesis has been supported in many experimental studies. For example, when people are randomly assigned to pursue a particular type of goal that mismatches their personalities, measured in terms of motive dispositions (Sheldon, Prentice, Halusic, & Schuler, 2015), intrinsic versus extrinsic values (Sheldon et al., 2019), narcissistic traits (Sheldon, Sedikides, Ntoumanis, Corcoran, & Titova, 2020), or agency vs communion orientations (Sheldon & Cooper, 2008), then they have measurably lower self-concordance (i.e. their goals are less organismically integrated; Sheldon & Kasser, 1995).

In sum, for the case of peoples' self-listed (idiographic) goals, self-concordance theory assumes and finds that feeling controlled or un-free in those goals is a symptom of having poor-fitting goals, and with (perhaps) having made unwise choices.

Why are people sometimes unable to make wise choices?

But how is it that poor choices can happen? From a deterministic perspective, shouldn't peoples' choices be directly dictated and determined by the needs of their organisms, with no fuss and bother? When a heart-rate increase is needed it is provided, by the sympathetic nervous system; when more blood glucose is needed it is provided, by the pancreas. But choosing selves often don't know what is required, or are at best guessing. That is, people seem to have the freedom to make thoughtless choices, to be ignorant of whom they really are, and to be badly confused about what is important and meaningful to them.

One way to think about how this can happen is in terms of the distinction between system 1 and system 2 functioning (Evans & Stanovich 2013; Kahneman, 2011; Sheldon, 2014). System 1 is the evolutionarily old behavioral system that operates automatically and mostly

non-consciously, continually providing us with perceptions and impulses, of which we may or may not take conscious notice. System 1, found in all vertebrates, is "fast," automatically synthesizing large quantities of information within broad impressions (Kahneman, 2011; Luoma & Martela, 2020). In contrast system 2 is verbal and cortical, operating linearly and logically. It is deliberate and "slow," as a mode of controlled processing. In an important sense, system 2 is always *last* on the scene -- although to an un-appreciated extent, its prior goal strivings can have been the *cause* of the scene that has now appeared (Sheldon & Goffredi, in press).

Explicit versus implicit motives research (Grund, Fries, & Rheinberg, 2018; Schultheiss, 2008) shows that system 2 often contains conceptual *beliefs* about the person's motives that are incongruent with their actual behavioral motives, down in system 1. A person's self-reported high need for achievement, for example, can be totally unrelated to that person's low achievement-focused behavior, measured objectively. Such incongruity can be very beneficial: System 2 can liberate humans from blindly following their non-conscious impulses, making it possible for them to override such tendencies (Stanovich, 2005). On the other hand, the fact that system 2 can misrepresent or fail to express the person's deeper wants and needs reveals a weakness of the system. Somewhat paradoxically, one indicator of our freedom is the fact that our intentional control systems are "free" to function poorly -- since they live up in the domain of concepts and ideas, which can be very disconnected from reality, including the reality of what is happening in the world, or what the person actually prefers. We can, for example, override our biological drive to eat, for various ideological reasons. Going on a hunger-strike may be very self-harming from an organismic point of view, but may be very noble from a morally principled point of view.

Another way to think about the existence of human free will is in terms of the evolved symbolic self that people create and live out within their minds, which functions as a behavioral executive within the person's action system (Sedikides & Skowronski, 1997, 2003). The symbolic self is our sense of playing a "social character" in the world (Sheldon, Gunz, & Schachtman, 2012), expressing ourselves to other selves and making choices in our lives. Symbolic self functioning corresponds closely to the activity of the *default mode network* in the brain (Andrews-Hanna, Smallwood, & Spreng, 2014; Sheldon, in press), whose primary function "is to support internal mental simulations that are used adaptively" (Buckner, 2013, p. 351). Although the symbolic self is "slow," it is not powerless; it has veto power ("free won't," the term coined by Benjamin Libet). The symbolic self can also make choices that take the person across the Rubicon, from deliberation to implementation (Gollwitzer, 2012). But the symbolic self is also able to say "not yet," and to ask its own mind further questions, which can provide further information, before decisions are finally taken. Like the CEO of a corporation, the symbolic self has the power to command much help and many internal resources, if it asks for them. Still, as mentioned above, it can also be partially cut off from its own deeper sources. As noted by the organismic theorist Ernst Angyal, "the relative segregation of the symbolic self within the organism is perhaps the most vulnerable point of the human personality organization" (Angyal, 1941, p. 171).

Countering three counterarguments to free will

In sum, we've laid out a way of thinking about free will such that the main problem isn't that we don't have it; rather, the problem is that we *do* have it, and may not know how to use it well (Fromm, 1941; Sartre, 1956). Our argument for free will builds on Christian List's approach

in focusing on intentional functioning, and adds the important consideration of whether peoples' intentions are wisely chosen. We assume that humans are ultimately unpredictable, but not because we behave randomly; rather, because we behave creatively, as a product of our highest-level mental processes operating within a rich symbolic field that contains nearly infinite degrees of freedom, limited only by our imaginations (which are not very limited, as our dreams show us).

The reductionist argument. Still, a reductionist might say: "Surely our mental intentions, suddenly emerging, aren't independent of the physical and brain processes that underlie them? As science improves, we will know more and more about how brains create complex thoughts, and someday, we will even be able to predict those thoughts, finally proving that people never had the freedom to do otherwise – they were always doomed to think and behave as they did."

We believe that reductionist arguments of this kind are ultimately unhelpful, because they represent the "naïve flat ontology" mentioned earlier, which says that the only true causal explanation of a phenomenon must be a lower-level physical explanation. Unfortunately, with this approach, there is no place to stop before reaching the bottom, i.e. the atomic level; no explanation above that level can be valid, because all explanations must be reducible, all the way down. A neuroscientist might say "the true cause of human behavior is always neurological, and subjective experience has nothing to do with it." But the neurological explanation is equally vulnerable to charges of being superfluous, given that the entire field of neuroscience might be explained away in lower-level (i.e. biochemical) terms, in the same way that subjective intentions might be explained away. But after 400 years of progress, surely we

don't need to abandon every science but physics? Rather than believing that all higher-level theories are ultimately redundant, it seems we need a layered ontology, which necessitates new predictive theories at every level.

The Predeterminism argument. A related argument against free will is *pre-determinism*, which says that whatever *did* happen, always *had* to happen. In this view humans are enmeshed in a giant machine whose cogs inexorably spit out the next moment, including our illusion of causing our own behavior. But this argument suffers from the same problem as the reductionist argument: that there is no place to stop. Logically, to embrace pre-determinism is to say that every event that has occurred in the 14 billion years of the universe was already set in stone, way back at the very first moment. You reading this sentence right now was always going to happen, no matter what! Whether or not this kind of determinism is true at the physical level is perhaps a matter of debate among physicists, but it is certainly not a useful perspective at the psychological or societal level, which are characterized by huge degrees of freedom for things to turn out differently.

We would again assert that there is no contradiction in saying that humans have free will and that the molecules in the human body obey the deterministic laws of causation. The two claims can be true at the same time. Free will is about *agential possibility,* while determinism is about *physical possibility,* thus operational at a completely different level (List, 2014, p. 161). The fact that we tend to see them contradicting each other is due to the naïve flat ontology where we fail to distinguish between these levels. The phenomena at the psychological level are multiply realizable in the sense that a myriad of different configurations of atoms can give rise to the same agential intention – the intention to grasp a coffee cup, for

example. And, building on a more inquiry-based ontology (Dewey, 1938), List notes that "our best guide to ontological questions in any domain is given by our best scientific theories of that domain" (List, 2014, p. 167). Our best theories at the atomistic level involve laws. Our best theories at the psychological and behavioral level involve intentions, beliefs, desires – and free will. Thus, our higher-level subjective intentions are best seen as a legitimate (but certainly not all-powerful) part of the determining formula of future events.

The epiphenomenalist argument. A third argument against free will comes from epiphenomenalism, which says that our experiences are mere byproducts of the real causes of our behavior, whatever those turn out to be. As the brain wetware does its thing it accumulates a "misty cloud of beliefs about itself" (Sheldon, in press), beliefs which can *affect* nothing, because they are nothing but *effects* -- like the smoke given off by a fire or the noise given off by a car. However it makes little sense that evolution would have given us the huge brain capacity required to experience ourselves as (seemingly) autonomous agents operating within a symbolic world, if that capacity did not provide some function or solve some problem. In Sedikides and Skowronski's (1997) model, the human capacity to create and animate a symbolic self evolved precisely for the purpose of helping us function within the incredibly complex narrative and social worlds of humans.

For epiphenomenalism to be an effective counterargument, it would need to be shown that human behavior could be better predicted without assuming any mental concepts such as intentionality, desires, or beliefs. If the lower level dynamics could explain all the mysteries of human behavior without referring to these higher-level concepts (and we doubt they ever will), then we could start taking epiphenomenalism more seriously, as we no longer would have any

need for the higher level concepts in explaining behavior. However, as long as peoples' intentions, desires, and choices feature in our best explanations of human behavior, it seems hyperbolic to call them incidental byproducts of the "real physical causes", which remain poor predictors, taken alone, of human behavior.

Some analogies. Returning to the hierarchical diagram, we believe the most accurate and scientifically fruitful account of human behavior is to say that human intentions operate at precisely the level of organization that you would expect them to: serving as the control processes that define and direct our broadest tactical decisions and initiatives -- with a causal force at their own level, that is not reducible to lower-level phenomena. Of course this doesn't mean we can't try to understand and better predict peoples' choices; but we will come closer to succeeding in this goal if we focus on psychological processes, rather than neurological or cellular processes.

An analogy or two may help. In order to explain what a computer does, moment to moment, we need to know more than just the physical architecture of the computer, its processor and memory, its bits and bytes; we need to know what's in the program that is making use of the computer, for the purposes of the program. The programmer and program control the actions of the computer, not the other way around; the hardware may set limits to what the software can achieve, but it's the software that provides the most predictive causal model of what the computer will do next. Similarly, the driver of a car controls where the car goes, not the car itself; the car merely supplies the cylinders, spark-plugs, and steering system.

In this view, both the computer and the car only provide *affordances*, ready to be put to work by control systems. In the human mind, the symbolic selves in which we live evolved to

utilize the affordances of our brains, by setting goals and regulating our action systems (Carver & Scheier, 1998; Sedikides & Skowronski, 1997). In our minds, we ask ourselves: "what do I want? Am I there yet?" These mental events can have irreducible influence on what happens next.

Does it really matter?

Does it really matter for us, personally, whether free will exists? We mostly feel free, whether or not we are in reality determined, in some scientific way. What if the question of scientific free will and the question of personal free will are really two different questions?

We suggest there are two major possibilities to consider. The first: that we are in reality determined, but we can still *feel* free, even if that feeling is in some sense an illusion. And, as SDT and research on lay beliefs in free will shows (Baumeister & Vohs, 2009; Crescioni et al., 2016; Stillman et al., 2010), we can derive better outcomes and more happiness in life, if we can cultivate more of that (somehow delusional) feeling of freedom. Note that even in this possibility, free will remains as real as, say, love, beliefs, desires, thinking, rationality, or any other concept operational at the psychological level. None of these things exist at the level where we examine human beings as heaps of molecules. Also note that concluding that they are all illusions from a physicalist point of view would not change much, as the physical level is anyways quite impotent in providing complete predictions about human behavior, and thus we would still need to continue to use these 'illusions' as our main way of interpreting and predicting things happening in our lives.

The second possibility is that we really do have "radical free will" (Sheldon, in press), but we may fail to recognize and use it – that is, we may sometimes feel controlled, when really, we

are always free. This second and more existential possibility (Sartre, 1956) is what we have argued for in this article, while suggesting that our feelings of being controlled are often a symptom of having chosen poorly, or of being in social environments that restrict our autonomy. In both cases the positive psychology prescription is the same: try to cultivate a greater feeling of freedom, of making autonomous and self-endorsed decisions. The only difference is whether these feelings of autonomy are ultimately just a (highly useful) illusion, or whether we should embrace the frightening possibility that we are actually in charge of our lives (Fromm, 1941), and should try to exert our will, as wisely as we can. You, reader, can freely decide which of these possibilities to accept!

References

Andrews-Hanna, Smallwood, & Spreng (2014). The default network and self-generated thought: component processes, dynamic control, and clinical relevance. *Annals of the New York Academy of Science, 1316*(1), 29-52.

Angyal, A. (1941). *Foundations for a science of personality*. New York, NY: Viking Compass.

Baumeister, R. F. (2008). Free will in scientific psychology. *Perspectives on Psychological Science*, *3*(1), 14–19.

Baumeister, R. F., Masicampo, E. J., & Vohs, K. D. (2011). Do Conscious Thoughts Cause Behavior? *Annual Review of Psychology*, *62*(1), 331–361.

Berlin, I. (1969). Two concepts of liberty. In *Four essays on liberty*. Clarendon Press.

Buckner, R. L. (2013). The brain's default network: Origins and implications for the study of psychosis. *Dialogues in Clinical Neuroscience, 15*(30), 351-358.

Carver, C. S. & Scheier, M. F. (1998). *On the self-regulation of behaviour*. Cambridge: Cambridge University Press.

Crescioni, A. W., Baumeister, R. F., Ainsworth, S. E., Ent, M., & Lambert, N. M. (2016). Subjective correlates and consequences of belief in free will. *Philosophical Psychology*, *29*(1), 41–63. https://doi.org/10.1080/09515089.2014.996285.

Deci, E. L. & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York: Plenum Press.

Dewey, J. (1938). *Logic—The Theory of Inquiry*. Henry Holt and Company.

Emmons, R. A. (1989). The personal striving approach to personality. In L. A. Pervin (Ed.), *Goal*

concepts in personality and social psychology (p 87-126). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Evans, J. St. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition advancing the debate. *Perspectives on Psychological Science*, *8*(3), 223–241.

Fromm, E. (1941). *Escape from freedom*. New York, NY: Farrar & Rinehart.

Fuller, M. D., Emrick, M. A., Sadilek, M., Scheuer, T., & Catterall, W. A. (2010). Molecular Mechanism of Calcium Channel Regulation in the Fight-or-Flight Response. *Science Signaling*, *3*(141), ra70–ra70. https://doi.org/10.1126/scisignal.2001152

Gollwitzer, P. M. (2012). *Mindset theory of action phases.* In P. A. M. Van Lange, A. W. Kruglanski, & E. T. Higgins (Eds.), *Handbook of theories of social psychology* (p. 526–545). Sage Publications Ltd. https://doi.org/10.4135/9781446249215.n26

Grund, A., Fries, S., & Rheinberg, F. (2018). Know your preferences: Self-regulation as need-congruent goal selection. *Review of General Psychology*, *22*(4), 437–451. https://doi.org/10.1037/gpr0000159

Kahneman, D. (2011). *Thinking, fast and slow*. New York, NY: Farrar, Straus and Giroux.

Kane, R. (2011). Introduction: The Contours of Contemporary Free-Will Debates (Part 2). In R. Kane (Ed.), *The Oxford Handbook of Free Will* (pp. 3–35). Oxford University Press.

Kim, J. (2002). The layered model: Metaphysical considerations. *Philosophical Explorations*, *5*(1), 2–20.

Klinger, E. (1977). *Meaning and void: Inner experience and the incentives in people's lives*. Minneapolis: University of Minnesota Press.

List, C. (2014). Free Will, Determinism, and the Possibility of Doing Otherwise. *Noûs*, *48*(1), 156–178. https://doi.org/10.1111/nous.12019

List, C. (2019). Levels: Descriptive, explanatory, and ontological. *Noûs*, *53*(4), 852–883.

Little, B. R. (2011). Personality science and the northern tilt: As positive as possible under the circumstances. In K. M. Sheldon, T. B. Kashdan, & M. F. Steger (Eds), *Designing positive psychology: Taking stock and moving forward.* New York: Oxford University Press.

Luoma, J., & Martela, F. (2020). A Dual-Processing View of Three Cognitive Strategies in Strategic Decision Making: Intuition, Analytic Reasoning, and Reframing. *Long Range Planning*, *Advance Online Publication*, 102065. https://doi.org/10.1016/j.lrp.2020.102065.

McAdams, D. P. (1996). Personality, modernity, and the storied self: A contemporary framework for studying persons. *Psychological Inquiry*, 7(4), 295–321.

McCrae, R. R. & Costa, P. T. (1996). Toward a new generation of personality theories: Theoretical contexts for the five-factor model. In J. S. Wiggins (Ed.), *The five-factor model of personality: Theoretical perspectives* (p. 51–87). Guilford Press.

Oppenheim, P., & Putnam, H. (1958). Unity of science as a working hypothesis. *Minnesota Studies in the Philosophy of Science*, *2*, 3–36.

Owens, D. (1989). Levels of explanation. *Mind*, *98*(389), 59–79.

Roberts, W. A. (2002). Are animals stuck in time? *Psychological Bulletin*, *128*(3), 473–489.

Ryan, R. M., Deci, E. L., Vansteenkiste, M., & Soenens, B. (in press). Building a science of motivated persons: Self-determination theory's empirical approach to human experience and the regulation of behavior. *Motivation Science*.

Sartre, J.-P. (1956). *Being and nothingness.* Philosophical Library.

Schultheiss, O. C. (2008). Implicit motives. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality psychology: Theory and research* (3rd ed., pp. 603-633). New York, NY: Guilford Press.

Sedikides, C., & Skowronski, J. J. (1997). The symbolic self in evolutionary context. *Personality and Social Psychology Review*, *1*(1), 80–102. https://doi.org/10.1207/s15327957pspr0101_6

Sedikides, C., & Skowronski, J. J. (2003). Evolution of the symbolic self: Issues and prospects. In M. R. Leary & J. P. Tangney (Eds.), *Handbook of self and identity* (p. 594–609). The Guilford Press.

Seligman, M. E. P., Railton, P., Baumeister, R. F., & Sripada, C. (2016). *Homo Prospectus*. Oxford University Press.

Seligman, M. E., Railton, P., Baumeister, R. F., & Sripada, C. (2013). Navigating into the future or driven by the past. *Perspectives on Psychological Science*, *8*(2), 119–141.

Sheldon, K. M. (2004). *Optimal human being: An integrated multi-level perspective*. New Jersey: Erlbaum.

Sheldon, K. M. (2011). Consilience in the biopsychosocial system. *Psychological Inquiry*, *22*, 52-65.

Sheldon, K. M. (2014). Becoming oneself: The central role of self-concordant goal selection. *Personality and Social Psychology Review*, *18*, 349-365

Sheldon, K. M. (in press). *Sailing the ghost ship: The case for radical free will*. Basic Books.

Sheldon, K. M. & Cooper, M. L. (2008). Goal striving within agentic and communal roles:

Functionally independent pathways to enhanced well-being. *Journal of Personality, 76,* 415-447.

Sheldon, K.M. & Elliot, A.J. (1999). Goal striving, need-satisfaction, and longitudinal well-being: The self-concordance model. *Journal of Personality and Social Psychology, 76*, 482-497.

Sheldon, K. M. & Goffredi, R. (in press). The self-concordance model: Using free will wisely. *Handbook of Research in Self-Determination theory*.

Sheldon, K. M., Gunz, A., & Schachtman, T. (2012). What does it mean to be in touch with oneself? Testing a social character model of self-congruence. *Self and Identity*, *11*, 51-70.

Sheldon, K.M., & Kasser, T. (1995). Coherence and congruence: Two aspects of personality integration. *Journal of Personality and Social Psychology, 68*, 531-543.

Sheldon, K. M., Prentice, M., Halusic, M., & Schuler, J. (2015). Matches between assigned goal-types and both implicit and explicit motive dispositions predict goal self-concordance. *Motivation and Emotion*, *39*, 335-343.

Sheldon, K.M., Prentice, M., & Osin, E. (2019). Rightly crossing the Rubicon: Evaluating goal self-concordance prior to selection helps people choose more intrinsic goals. *Journal of Research in Personality*, 79, 119-129.

Sheldon, K.M., Sedikides, C., Ntoumanis, N., Corcoran, M., & Titova, L. (2020). Narcissism and social motives: Successful pursuit of egosystem goals boosts narcissism. *Self and Identity,* 19:7, 841-862.

Stanovich, K. E. (2005). *The robot's rebellion: Finding meaning in the age of Darwin*. University of Chicago press.

Stillman, T. F., Baumeister, R. F., Vohs, K. D., Lambert, N. M., Fincham, F. D., & Brewer, L. E. (2010). Personal Philosophy and Personnel Achievement: Belief in Free Will Predicts Better Job Performance. *Social Psychological and Personality Science*, *1*(1), 43–50.