**Bringing self-control into the future**

Samuel Murray, *Providence College*

***Abstract***

The standard story about self-control states that self-control is limited, aversive, and that the function of self-control is to resist impulses or temptation. Several cases are provided that challenge this standard story. An alternative, future-oriented account of self-control is defended, where the function of self-control is to manage interference that arises from overlapping information processing pathways. This provides a computationally tractable account of self-control rooted in one’s being vigilant. Self-control manifests the maintenance dimension of vigilance. This not only provides an attractive alternative to the standard story about self-control, but also indicates how the *future-oriented* aspects of self-control are more fundamental than the *present-oriented* aspects.

*Introduction*

The standard story about self-control centers around three claims. First, many think that self-control is limited. Exercising self-control at one moment makes it more difficult to exercise self-control thereafter in the absence of a recovery period. There are a variety of explanations for these limitations, ranging from consumption of limited resources (Baumeister et al., 1998) to structural and functional limitations of processes supporting self-control (Edin et al., 2009; Ma & Huang, 2009; Miller, 1956; Oberauer & Kliegel, 2006; Usher, Cohen, Haarmann, & Horn, 2001). Second, exercising self-control requires effort and is, for that reason, aversive (Kool et al., 2010; Kurzban, 2016; Inzlicht, Shenhav, & Olivola, 2018). Finally, the function of self-control is to resist temptation or suppress impulses (Ainslie & Haslam, 1992; Bargh & Chartrand, 1999; Holton, 2009; Joosten et al., 2015; Mischel et al., 1989; Muraven & Baumeister, 2000; Myrseth & Fishbach, 2009; Soutschek et al., 2016; Watson, 1977).

Discussions of self-control are historically tied to explanations of weakness of will (Mele, 2011). Framed in this way, the standard story of self-control makes sense. In such a case, the agent must employ self-control to avoid succumbing to temptation (functional claim). Because of the temptation, the agent must use self-control to resist something she currently desires (aversion claim). Lastly, if the individual does not remove herself from the tempting scenario, then she eventually gives in because her self-control is depleted (limitation claim). It’s a nice package. And it’s incorrect.

We can challenge the standard story from the bottom up. While self-control is sometimes employed to resist temptation or suppress impulses, this certainly does not exhaust the range of applications for self-control. To see this, consider the following example:

*Hot Car*. Kate normally dropped her daughter Sam off at daycare before working from home. But one summer morning, Kate had an early meeting and her daughter was sleeping in. This meant Brad—Kate’s husband—had to take Sam in before heading to work. Unfortunately, Brad forgot to drop his daughter Sam off. By the time he remembered, the little girl had been sitting inside a car, windows closed, on a summer day in Texas for several hours. EMT’s declared Sam dead a mere 80 minutes after Brad called 9-1-1.

This, I submit, is a failure of self-control. But where’s the temptation? Where’s the impulse? You won’t find them because they aren’t there. This suggests that there is more to self-control then the motivational dimension encapsulated in the functional claim of the standard story.

Self-control is not exclusively aversive. Theorists typically infer aversion from the fact that exercising self-control requires effort. But neither self-control nor effort are necessarily aversive. Consider the fact that learning a new language, or a new instrument, requires lots of self-control and effort, but neither experience is aversive (or, at least, not always aversive). The aversion claim ignores a general truth, namely that some of the enjoyable experiences in life *also* require lots of effort.[[1]](#footnote-1)

Discussions of limitations on self-control are complicated, and we should distinguish the target of the discussion. There is the phenomenon of self-control limits and the candidate explanations of those limits. Let’s consider first the phenomenon of self-control limitations. In a standard study on self-control limits, participants in the experimental condition perform two tasks, both of which require self-control. The control group performs a non-control task in the first block and a self-control task in the second block. These tasks range from Stroop tasks to emotion suppression tasks. Participants in the experimental condition typically perform worse on the second task (or tasks following an initial self-control task) across a variety of dimensions relative to the control group (Schmeichel, 2007; see Hagger et al., 2010 for a meta-analysis of depletion effects in self-control studies). Some of the findings include lower inhibition after performing a controlled task (Vohs & Faber, 2007), stronger bias toward behaviors that require only automatic processing (Schmeichel, Vohs, & Baumeister, 2003), higher rates and intensity of prejudicial judgments (Muraven, 2003), and diminished performance on tasks that require a speed/accuracy tradeoff (DeWall, Baumeister, Mead, & Vohs, 2011).

These depletion effects have shown up in hundreds of independent studies (see Baumeister & Vohs, 2016: 75-93 for a summary of findings). A meta-analysis showed that effects from these various studies were both robust and significant (Hagger et al., 2010). However, some challenge whether this meta-analysis appropriately accounted for possible publication and small-study biases (Carter & McCullough, 2014). Following up on this possibility, a later meta-analysis found a corrected effect that is effectively zero (Carter et al., 2015). And, though a subsequent preregistered, multi-lab replication effort found significant effects (Sripada et al., 2016), a meta-analysis of the results found an ego-depletion effect close to zero (Hagger et al., 2016), and Bayesian analyses of the results supported this null finding (Etherton et al., 2018). Additionally, attempts to replicate some of the training effects on ego depletion found in the original Hagger et al. (2010) meta-analysis failed when correcting for timescale and task type (Miles et al., 2016). When participants performed controlled tasks that lasted over longer timescales (i.e., tasks with higher ecological validity), Wenzel et al. (2019) found a reversed ego depletion effect.

What this means is that ego depletion effects are more complicated than original two-task experimental conditions suggested. While there is some evidence that ego depletion effects do not exist, it seems more probable that ego depletion effects are highly sensitive to contextual factors. This would explain the highly variable findings. In fact, in what is becoming a near weekly tradition, yet *another* meta-analysis of ego depletion studies found significant overall effects of depletion across nine different task types (Dang, 2018). Hence, it seems safe to say that ego depletion effects are real, so we need some explanation for them.

One explanation is that exercising self-control consumes a limited resource. Hence, an exercise of self-control depletes the agent’s resources. When this occurs, a central governing mechanism registers that the agent is consuming resources at an unsustainable rate, thus causing the brain to reduce exertion and conserve energy (see Evans, Boggero, & Segerstrom, 2016). This framework construes self-control as analogous to a muscle and willpower as analogous to physical strength and posits that the limited resource is metabolic (see Baumeister & Vohs, 2016).

There are two problems with this view. The first is that the Strength Theory has difficulty explaining the variety of contextual factors that modulate ego depletion effects (see Kurzban et al., 2013). The second is that there appears to be no plausible metabolic substrate that supports self-control. The original proposal for the metabolic substrate was glucose (Gailliot & Baumeister, 2007 and Gailliot et al., 2007). This proposal has not fared well over time. Later studies found that exerting self-control sometimes *raised* blood glucose levels over time (Baumeister & Vohs, 2016). A meta-analysis of various glucose-related results found no significant effects (Dang, 2018). The proposal is also highly counterintuitive. Brain processing consumes a miniscule amount of glucose relative to the rest of the body. Exercising self-control raises glucose consumption by a miniscule fraction. If such small increases in glucose consumption are sufficient to trigger depletion effects, then self-control would be much too volatile to support temporally extended agency (Wenzel et al., 2019).[[2]](#footnote-2)

In the wake of suspicion surrounding the Strength Theory, several alternative theories about the nature of self-control have emerged. Many of these alternatives cite motivation and attention as essential components of self-control (e.g., Botvinick & Braver, 2015; Eastwood, Frischen, Fenske, and Smilek, 2012; Inzlicht, Schmeichel, & Macrae, 2014; Hockey, 2011; Inzlicht, Shenhav, Olivola, 2018; Kool & Botvinick, 2014; Kurzban et al., 2013; Shenhav et al., 2017). There are some important differences between these theories, but the central thread connecting them is that ego depletion effects result from shifts in motivation (thereby causing shifts in attention). For example, Inzlicht et al. (2014) explain that one explanation of ego depletion effects in laboratory studies is that participants are motivated to think about other things, like what they’ll be eating for dinner later or what they’ll do with their friends over the weekend. This also explains why shifting incentives mitigates the ego depletion effect. Limitations of self-control, then, reflect our fleeting motivations to carry out any single task (and, perhaps, our underlying propensity to shift attention from tasks we must do to tasks we want to do).

While the Motivational Theory avoids the problems that plague the Strength Theory, the view is also unintuitive in some respects. Consider, again, the *Hot Car* case. There, we identified a self-control failure, but it’s not clear that there’s a shift in motivation that explains the self-control failure. In fact, we can press this point and say that it’s *wildly implausible* that parents in these kinds of scenarios have changed their preferences or shifted motivation. That would imply that the parent has come to value getting to work on time over the life of their child. We see this inadequacy when we extrapolate out to other performance breakdowns. For example, forgetful cooks do not suddenly lose motivation to have their kitchen not be engulfed in flames. And forgetting a birthday or anniversary does not reflect an individual’s devaluation of a relationship.

Both major theoretical approaches to self-control turn out to be inadequate. Thus, we need some approach that explains apparent limitations on self-control (and suggests some principled explanation of the influence of contextual factors on these limitations) in an empirically tractable way that maps onto the various dimensions of self-control. Additionally, it would be nice for the theory to explain the connection between effort and self-control in a way that avoids construing self-control as necessarily aversive *and* to explain where temptation resistance fits into the scheme of self-control. Finally, we want our resulting theory of self-control to explain what’s going wrong in the *Hot Car* case.

In this chapter, I propose a theory of *self-control as vigilance*. In Section 1, I explain why we should expect vigilance and self-control to be related. In Section 2, I propose an alternative way to understand limitations on control that grounds self-control limits in efficient cognitive network architecture. Section 3 discusses the computational principles that govern control allocation. With this in hand, in Section 4 I outline the maintenance dimension of vigilance and discuss two modes of maintenance, portrayal and recovery, and the relationship between these modes and the computational account of control allocation from Section 3. Finally, Section 5 ties together various threads, discussing the connection between self-control and effort, the kind of self-control that temptation resistance requires, and the explanation of various ego depletion effects.

1. *Self-control as vigilance (I)*

To see the reasons for connecting self-control to vigilance, let’s step back and consider what the purpose of self-control is. If we reject the claim that the function of self-control is just to resist temptation or impulse, then why do creatures like us have self-control at all?

Agents like us pursue a variety of goals and commit to a wide range of projects. Living a meaningful life requires heterogeneity of pursuits. However, these goals cannot be simultaneously realized or even simultaneously pursued. Thus, multiplicity of goals, combined with limits on cognitive and bodily resources, generates the need for a goal maintenance mechanism (Murray, 2020). Vigilance regulates goal-relevant information and the implementation of goal-relevant information in mind that facilitates goal-directed action (Murray and Vargas, 2020). A perfectly virtuous creature with heterogeneous pursuits and certain cognitive limitations that experiences no temptations and is subject to no impulses would still require vigilance for effective goal maintenance. Further, it seems intuitive to think of part of the function of self-control as goal maintenance. Self-control, then, amounts to the pursuit of multiple goals in an effective and efficient manner.

One benefit of this account is that it classifies obsessions as failures of self-control. This is not possible on the standard story of self-control. The obsessed individual does not succumb to temptation or impulse; rather, she focuses too much on one thing at the expense of everything else. Obsession, then, manifests lack of efficiency and, hence, lack of self-control to some degree.

Given the overlap between the functional description of vigilance and the functional description of self-control, it makes intuitive sense to consider the claim that self-control is a species of vigilance.[[3]](#footnote-3)

2. *The limits of control*

Why is it that an Apple Watch—with four circuits—can calculate 2 two-digit math problems simultaneously, but the human brain—with over 20 billion cortical neurons—*cannot*? As the question suggests, control limitations derive neither from metabolic limitations (extra sugar won’t help) nor from structural/functional limitations (as a structurally and functionally limited computational device can do things the human brain cannot).

The Stroop task provides another example on the limits of control. Try as hard as you might, you cannot engage in color identification and word naming simultaneously (if you could, incongruent Stroop tasks wouldn’t be harder and take longer than their congruent counterparts). You cannot imagine what it’s like to be in Berlin *and* what it’s like to be Bogotá simultaneously. The examples are endless. But, again, it does not seem like these limitations are ultimately reducible to structural or functional limitations. The human brain has an enormous number of resources that it *could* devote to controlled processing. So, even if structural or functional limitations are part of the explanation for limited controlled processing, it seems that something else is fundamentally explanatory.

But, if we’re not going to appeal to metabolic, structural, or functional limitations, then what might explain these strict limitations on control? To get to the answer, consider an example. We utilize task representations to guide goal-directed behavior. So, we store task representations that correspond to crossing the street, whisking the eggs, and watering the garden. However, you wouldn’t want a unique representation for *every* unique task you can perform. For example, you don’t need a separate street-crossing task representation for every street you happen to cross. A single, generalized ‘street-crossing’ task representation will do (see Rougier et al., 2005).

This is true in general. Representations that can be utilized for multiple tasks are more efficient than highly specialized representations. Drawing from a basic stock of generalized representations that can be flexibly deployed across various task types is known as *multiplexing*. Allport, Antonis, & Reynolds (1972) were the first to suggest that the brain multiplexes. Multiplexing, however, introduces the possibility of channel crosstalk. If you have a variety of available input-output task mappings subserved by the same representation, then these tasks might potentially interfere with each other (Forbus, Gettner, & Law, 1995; Hinton, McClelland, & Rumelhart, 1986). The Stroop task provides a simple example. The input-output mappings that correspond to word naming and color identification both utilize the same representation. Hence, the processes that subserve these two tasks cannot be activated simultaneously and activating one implies performance deficits for the other. Thus, multiplexing correlates negatively with the capacity to multitask, as the more multiplexing occurs, the greater the possibility of crosstalk, which reduces the capacity to multitask without possibility of interference (see Feng et al., 2014 for a computational model of the absolute limits of multiplexing on multitasking in an optimal control network independent of network size and number of control nodes).

Dramatic limitations on control reflect the brain’s preference for efficient coding through multiplexing. We can infer the brain’s preference for multiplexing over multitasking from the fact that we can perform so few controlled tasks simultaneously. This, however, generates a problem of interference. A high degree of pathway overlap implies several potential sites of interference between task mappings. To solve this problem of interference, you need a control manager. Thus, Jon Cohen claims that: “These [shared resource] models suggest that constraints on the simultaneous execution of multiple tasks can be viewed as the *purpose* of control, rather than a limitation in its ability” (2017: 5, emphasis original).

The function of the manager unit, as Cohen suggests, is to adjudicate conflicts between overlapping pathways and bias lower-level information processing in ways that support goal-directed behavior (Botvinick et al., 2001). Biasing is accomplished by allocating control to unique pathways to alter the threshold for a neuronal population to fire. Hence, conflict management requires allocation of control (Section 3 discusses the computational principles of control allocation).

One natural question is why the brain (or the evolutionary principles that govern brain development) did not generate unique control units for every site of potential crosstalk. However, the addition of extra control units (beyond the optimal convergence point) generates performance *deficits* (see Feng et al., 2014: Fig. S5). Both error rate and expected reward attainment decrease in proportion to additional control units. Feng and colleagues do not offer a computational story for the performance deficits, but the effect makes intuitive sense. Efficiency declines when there are too many cooks. There might also be adaptive value in having fewer control units. With fewer control units, local and global reconfigurations of task mappings in response to task acquisition can occur more easily. This is because a system that utilizes minimal control units requires a simpler network architecture. Over time, this might represent a metabolic advantage over massively modular architectures (Anderson, 2014: 38).

This framework provides a new way to think about limitations on self-control. Multiplexing, combined with a small number of control nodes, constrains the capacity to multitask. This computational account of control limitations is more empirically tractable than the metabolic account and explains the structural and functional limits on controlled processing. Return to the Apple Watch. The watch doesn’t multiplex, instead utilizing discrete informational units for each calculation. This is fine for the watch because it doesn’t do much and so can afford to prefer multitasking. We, however, have the advantage of being able to do innumerable things. The price for this is that we multiplex for the sake of efficiency. This generates pathway overlap that decreases capacity to multitask. However, we can solve this interference problem with control nodes that manage channel interference. This generates a new question: what are the computational principles that dictate allocation of control across these various channels?

3. *The expected value of control*

One cornerstone of cognitive architecture research is that cognitive efficiency is built on a speed/accuracy tradeoff (Bogacz et al., 2006). There are benefits associated with a system of heuristics and defaults (e.g., fast processing speed), but there are also costs (e.g., error and inflexibility) (Miller & Cohen, 2001). There are benefits associated with a deliberative system that assesses available actions relative to an internal causal model of the environment (e.g., accuracy and flexibility), and there are costs (e.g., slow processing speed).

The computational principles that govern control allocation should, then, solve for an optimal balance between speed and accuracy. A background presumption is that the network is configured for long-term maximization of reward. Thus, the optimal balance between speed and accuracy reflects an aim toward maximizing reward. But, given the discussion from Section 2, the optimal balance must be measured relative to the bounded computational powers of a control network that utilizes shared representations. One engineering principle that seems intuitive is that the network should rely on a system of default settings, with control units intervening only when necessary. This is plausible because optimal network design would suggest that you use the fast/computationally cheap system as much as possible and correct for errors (with the slower/computationally expensive system) when necessary. Control units, then, would monitor for conflicts among lower-level information processing units. When conflict is detected, the control unit would calculate the costs and benefits of engaging control to resolve conflict.

Recent computational models of control allocation provide a framework for thinking about these issues at the psychological level. That is, if we understand the computational principles that govern control allocation at the neurobiological level, we will gain a foothold for understanding how those principles manifest at the psychological level. One popular model is the Expected Value of Control (EVC) model articulated by Jonathan Cohen (see Shenhav, Botvinick, & Cohen, 2013 for an early statement of the computational and mechanistic aspects of the view). The mechanistic aspects of the EVC are not important here, so I will focus only on the computational elements. The following three computations represent the core of this evaluation (these equations are taken directly from Shenhav, Botvinick, and Cohen, 2013: 221):

***Equation 1***:

EVC(signal, state) = [ ∑*i Pr*(outcomei|signal, state) • Value(outcomei)] – Cost(signal)

***Equation 2***:

Value(outcome) = ImmediateReward(outcome) + γmaxi[EVC(signali, outcome)]

***Equation 3***:

signal\* ← maxi[EVC(signali, state)]

The computational principle of the EVC calculates the value of a particular control signal based on three components. The first component represents the probability of achieving a particular outcome given that a control signal is sent in a particular environment (*Pr*(outcomei|signal, state)). The second component is the value of the outcome (Value(outcomei)), and the third component represents the intrinsic cost of the signal (Cost(signal)).

Currently, there are no widely accepted theories of which mechanisms compute values for the probability of achieving an outcome given transmission of a control signal nor what computational principles govern these mechanisms. Some suggest that certain mechanisms simulate controlled behavior to rapidly generate estimates (Pezzulo, Rigoli, & Chersi, 2013). Others suggest that model-free learning mechanisms update estimates of probabilities for the system without the use of simulation (Gershman, Horvitz, & Tenenbaum, 2015; Braem, 2017). Others find evidence for the use of heuristics to estimate probabilities without assessing task-specific demands (Dunn, Lutes, & Risko, 2016). At this point, not enough is known about task representation acquisition and updating to know *how* this process is carried out. But the point remains that the computations underlying control allocation are sensitive to the probability of achieving the outcome. When the probability of success diminishes, control decreases (Kool et al., 2017-a).

The representation of value has two parts. The first concerns the immediate expected reward upon achieving the outcome. Part of the cost-benefit analysis associated with determining control allocation should be sensitive to the potential *benefits* of achieving the outcome. However, the EVC model includes a temporal component to value representation. Outcome value representation reflects the kinds of action-reward sets that will be available from the achieved outcome state (γmaxi[EVC(signali, outcome)). Thus, a highly immediately rewarding outcome will diminish in overall value if there are no desirable action-reward sets available from the state in which one has attained the outcome.

This aspect of the value representation reflects part of the cost of control (but not the intrinsic cost represented in the third component). Consider two possible outcomes, O1 and O2. Suppose that the immediate reward associated with these outcomes is such that O1 > O2. But, achieving O1 would put you in a position where you could not pursue other rewards without some ‘rest’ (the quotes suggest resisting an interpretation of limited resource consumption; the issue of limitations is discussed below). Achieving O2, on the other hand, puts you in a better relative position to achieve further rewards. Of course, the value of O1 might greatly exceed O2, thereby nullifying these costs. The point is simply that the computation is sensitive not just to the value of available outcomes, but about future deployments of control.

Finally, there is the cost of the signal. Recently, there has been debate about how to interpret the cost of control. Roughly, this represents the *intrinsic* costs of control. Recall that our cognitive network architecture is designed so that only one (or a very small number) of controlled tasks can be pursued at any one time. When we utilize control, we intervene on the ‘default’ settings of lower-level information processing pathways, thereby configuring the system to support goal-directed behavior in the circumstances. However, this configuration means that any number of available configurations are not available, meaning that the rewards associated with pursuing other actions are missed. Hence, we can understand the intrinsic costs of control in terms of the opportunity costs associated with not pursuing other goals.

It is worth noting that there are two separate cost values computed. The first is part of the outcome value representation. This cost reflects ‘depleted’ resources that result from using control. This depletion, however, is metaphoric. The real cost consists in task-switching costs associated with reconfiguring the system either toward another goal or back to a default (unconfigured) state. This also implies diminished behavioral flexibility relative to non-goal-related activities, thereby increasing switching costs. The intrinsic cost, on the other hand, reflects the opportunity cost of thinking and pursuing one goal at the exclusion of others or at the exclusion of plural (unconfigured) goal pursuit.

The EVC model assigns a unique value to a particular control signal as a function of the probability of achieving an outcome given the transmission of a control signal, the value of the outcome, and the cost of that signal. To see how the model works, consider an idealized, non-iterated incongruent Stroop task condition. Here, one is presented with a word (e.g., ‘green’) filled in with a different color (e.g., RED) and told to identify the color. When the word is presented, there is a conflict among lower-level information processing units associated with word naming and color identification (since both utilize the same representation). The EVC model describes the calculations that control units compute to determine whether to allocate control. Because the scenario is idealized, we can assume that the probability of achieving the outcome given the signal = 1, while the probability of achieving the outcome without the signal = 0. The value of the outcome-given-signal is high (as it conforms to experimenter instructions), so we can assign it a value of 1, whereas the value of outcome-without-signal is 0. The cost of the signal drops out here, as the Stroop task in this case is not repeated. Hence, relative to the experimental condition, there is no need to calculate values relative to the state where one has achieved the outcome. The expected value of the control signal outweighs both its costs *and* the alternative based on not signaling. With the computation complete, the control unit signals to bias the network toward goal-relevant processing. Of course, the computations get more complicated as probabilities are added, tasks become temporally extended, and the value of other control signals must be weighed. But the idealized example shows what the various components are meant to represent.

The EVC model is one of many available models of the cost-benefit analysis associated with control allocation. So why go on at length about the components and implications of this one model? There are good reasons to prefer this model to others. For one, the EVC model has plausible neurobiological realizers and provides an integrative framework for thinking about various neurobiological mechanisms (Shenhav et al., 2017). Recent work seems to show that we can map EVC computations to various functionally localized neural mechanisms (Shenhav, Botvinick, & Cohen, 2013 for a review). Also, experimental work supports the EVC model, especially in reinforcement learning paradigms, where the EVC makes divergent predictions from other control allocation algorithms (e.g., Kool et al., 2017-b). Finally, the model connects with recent theoretical advances in artificial intelligence research and machine learning applications. For example, the EVC model mirrors theoretical frameworks that model rational metareasoning and algorithm selection in artificial intelligence systems (Lieder et al., 2018) and implementing the EVC model has produced advances in machine learning (LeCun, Bengio, & Hinton, 2015). In addition, the EVC model seems compatible with other computational models of control allocation (Musslick and Cohen, 2021), but that argument lies outside the scope of the present chapter. Should the EVC model turn out incorrect, it is likely that whatever the correct model is will be close enough to the EVC that the applications to vigilance remain apt.

4. *The Maintenance Dimension of Vigilance*

In this section, I will outline the connections between the EVC model and vigilance. The maintenance dimension of vigilance maps most directly onto the various components of the EVC model (though I’ll also argue that implementation corresponds to the EVC, too, so that the EVC model provides a fully computational account of vigilance). This will then give us a template for discussing various failures of vigilance. In the next section, I’ll discuss the relationship between EVC, vigilance, and self-control.

First, I will briefly outline an account of vigilance I have defended elsewhere (Murray, 2017; Murray, 2020; Murray and Vargas, 2020; Murray and De Brigard, 2021). Being vigilant is realized in virtue of three mental events that facilitate plan-directed awareness: (1) monitoring for circumstantial and task-relevant information that, when perceived (2) triggers implementing task-appropriate representations that are (3) maintained through the completion of the task (or task-segment) or until the agent revises their intention. Being vigilant is realized whenever these operations aim to produce task-congruent temporally extended action. These operations correspond to two different dimensions of vigilance: monitoring and maintenance.[[4]](#footnote-4)

These dimensions can issue in distinctive kinds of performance breakdowns whenever some operation fails to occur. There are failures to recall a task set, failures to preserve a task set, failures to coordinate two tasks, and failures to coordinate tasks with known information. In the former two types of failures, the agent fails to apprehend a presently available consideration that is relevant given some previously adopted plan. In the latter two types of failures, there is a failure to coordinate between one’s present and future self. In other words, there is a failure to apprehend that some future consideration is relevant to one’s current planning and task set acquisition. Thus, in the former cases the agent fails to apprehend that a *present* consideration is relevant given *prior* planning, whereas in the latter cases the agent fails to apprehend that a *future* consideration is relevant to one’s *current* planning. All share the common feature of exhibiting a failure to move from an adopted plan to relevant features of an action context.

Each failure has a different temporal orientation (past or future). Both dimensions of vigilance have temporally variegated deployments. Thus, the four kinds of failures correspond to the four different aspects of vigilance (present monitoring/maintenance and future monitoring/maintenance).Here, I want to focus on the maintenance dimension of vigilance. I call the future-directed aspect of maintenance *portrayal*, while the present-directed aspect of maintenance is called *recovery*. To get a sense for these different aspects, let’s again look at what happens when they breakdown.

In the Hot Car example, the parent plans to drop the kid off at school before heading to work but ends up going straight to work. This is a present-directed case of maintenance failure (hence, a failure of recovery). The parent fails to adequately respond to goal threats and goal substitution. What happens is that, in general, favoring task sets that minimize the cost of control leads to reliance on habitual routines. The parent, when putting the kid in the backseat, fails to accurately assess the need for engaging in controlled processing for goal maintenance. In one sense, this is understandable. As we saw in the previous section, there is a tendency toward minimizing the use of control. Given the past record of success when relying on habits, there seems to be no need to deliberate about the use of control, much less initiate an actual deployment of control. From the parent’s perspective, both the likelihood of failure and the task demands are so low that there seems to be no need to think very hard about the task of dropping the kid off.

This can be represented in terms borrowed from the EVC model. Recall that there are three representational components to the model: the probability of achieving an outcome given a signal, the value of the outcome, and the cost of the signal. Misrepresentations could occur at each point. The parent might misrepresent the probability of achieving the outcome in the absence of a control signal, she might misrepresent some aspect of the value of the outcome (though this is unlikely), or she might misrepresent the cost of the control signal. As the case is described here, it seems that the parent misrepresents the cost of control. Rather than move from a default to controlled configuration, the parent instead sees no need to pay the cost of control.

This might seem fishy. You might think: “Wait, the *parent* is misrepresenting? Surely people don’t perform these computations at the personal level. These *misrepresentations* are subpersonal glitches, not anything you can pin on the agent.” The answer to this objection, however, provides a bridge between the agent and the computational principles outlined in the previous sections. Roughly, the computational failure can be attributed, in failure of vigilance cases, to the agent’s failure to construe her tasks appropriately (in other words, the agent fails to structure her plans appropriately, leaving those plans susceptible to goal threat or substitution).

The way that an agent construes her goals is important. Literature on implementation intentions suggests that concretely structured plans are less likely to be abandoned prematurely (Gollwitzer, 1999). Similarly, task construal can have an impact on how attractive some tempting item seems (Fujita et al., 2006). Task construal bears on plan structure, as task construal typically implies a certain plan structure (Ho et al., 2022). For instance, there’s enormous difference between forming a plan to exercise as ‘I’ll workout when I have the time’ and ‘I’ll only take a rest if I’ve worked out the previous two days’ (Fujita, 2011). Similarly, recovering alcoholics who adopt a narrower, more present-directed focus of their circumstances (low-level construal) tend to relapse at higher rates than recovering alcoholics who adopt a more global perspective (high-level construal (Keough, Zimbardo, & Boyd, 1999).[[5]](#footnote-5)

This might make it seem like all failures of vigilance are just failures of construal, thereby implying that there is just one kind of failure of vigilance, not four. With the EVC in hand, we can outline more precisely the four different kinds of vigilance failure. This account depends on the functional distinctness of maintenance and monitoring.

One speculative difference between monitoring and maintenance concerns the calculation of probability. Recall, from Section 3, that separate mechanisms calculate probability estimates for achieving an outcome. Perhaps the mechanism of monitoring consists in these probability calculators, whereas the mechanism of maintenance consists in cost calculators (value representations would be subserved by valuative—possibly reinforcement-learning—mechanisms). Hence, the aim of monitoring at the information-processing level is probability calculation, while the aim of maintenance at the same level is cost calculation.

This connection to the EVC suggests the following taxonomy of vigilance failures. We can distinguish broadly between two kinds of failures: maintenance failures and monitoring failures. Within each type, there can be either present-directed or future-directed failures. These four failures can be characterized in psychological terms. A present-directed maintenance failure consists in failing to preserve a task set over time. A future-directed maintenance failure consists in a failure to coordinate multiple task sets over time. A present-directed monitoring failure consists in failure to recall a task set at the appropriate time. And a future-directed monitoring failure consists in “double booking”, or committing to practically incompatible tasks. While each of these failures can be characterized at the psychological level, each failure can also be characterized at the computational level based on elements of the EVC. A present-directed maintenance failure consists in inaccurately representing the cost of a control signal. A future-directed maintenance failure, on the other hand, consists in inaccurately representing the *future* costs of some present course of action. A present-directed monitoring failure consists in an incomplete representation of the decision space, while a future-directed monitoring failure consists in inaccurately representing the probability of success contingent on sending a control signal.

The case of present-directed monitoring failures requires some discussion. Part of the activity of monitoring is to accurately represent features of the action context relevant to one’s planning. At the computational level, this amounts to accurately entering values for the various arguments in the EVC function. For example, suppose I’m telling a funny story around you that recounts your doing something embarrassing. I don’t even recognize that you’re getting upset and I continue telling the story. Here, I’ve misinterpreted the situation such that certain values (like your taking offense at the anecdote) are not even represented in my decision procedure to continue telling the story. This seems to be a failure of monitoring, but on a slightly different order than other kinds of failures of vigilance.

Note that each aspect of vigilance maps to a component of the EVC. This strengthens the claim made earlier that the EVC provides a fully computational account of vigilance. Additionally, the EVC model leaves room for the importance of task construal. The representation of plan structure might play an important role at the computational level. The foregoing at least suggests that failures of monitoring and maintenance are distinct, even though they both share a common constituent, namely task construal. This also demonstrates the dynamic interactions between the different dimensions of vigilance. For instance, the representation of cost might be sensitive to representations of probability and vice versa (consider the Hot Car case; perhaps the representation of probabilities affects the representation of current costs of control. Not enough is currently known about the relations between these different aspects of the computations, so nothing definitive can be stated about dependence or fundamentality of one relative to the other).

The EVC model provides a computational underpinning for the functional account of vigilance. The reason for connecting these two constructs is not to reduce vigilance to the EVC model. Of course, an individual that realized a psychology that failed to be accurately described by either the computational or representational theory of mind might have vigilance. The point of tying the vigilance account to the EVC model is that it provides a computationally tractable way of thinking about the activity of vigilance. When tied to a sufficient mechanistic model of vigilance, EVC enables predictions about when and where vigilance will breakdown and how we might improve the exercise of vigilance.

5. *Self-control as vigilance (II)*

What remains is to determine whether the account of vigilance and cognitive control adequately captures certain features of the self-control construct. In this section, I’ll show how vigilance explains the activity of various processes typically associated with self-control. Additionally, I’ll consider two potential problems with the self-control as vigilance theory, namely that there are empirical reasons to dissociate cognitive control and self-control and that the function of self-control is exhausted by temptation resistance.

There are three foundational components typically associated with models of self-control: inhibition, task updating, and task switching (Davisson & Hoyle, 2017). Additionally, there are four components typically associated with cognitive control: working memory, response selection, response inhibition, and task-set switching (e.g., Sabb et al., 2008). The EVC model makes explicit connection between cognitive control and vigilance. To connect this to components of self-control, consider that vigilance structures the flow of information into working memory. The activity of monitoring aims to bring plan-relevant information to mind at the appropriate time, while filtering out information when it is no longer relevant. Bringing information to mind here amounts to bringing the information into working memory (Baddeley, 2012). This shows that updating is a component of vigilance and, by extension, that cognitive control is partly constitutive of self-control. As for inhibition and switching, these components fall under the maintenance dimensions of vigilance. Sending a control signal (which is constitutive of the activity of vigilance) is constitutively relevant to inhibiting some response for the sake of selecting some other response.[[6]](#footnote-6) Similarly, when the cost of some ongoing controlled activity becomes too great (relative to other goals that one could pursue), then the agent shifts. Again, this is just the function of the maintenance dimension of vigilance in its present-directed aspect. Thus, it seems possible to map both cognitive control and self-control onto the functional account of vigilance, thereby linking all three constructs.

One worry about this account is that it seems to make self-control back into mere impulse control/temptation resistance. After all, if conflict monitoring is part of the maintenance dimension of vigilance (which dimension, recall, is the one associated with self-control), then isn’t self-control still just a form of conflict management? But the objection gets the view wrong. Conflict monitoring *may* involve detection of distractions, temptations, or impulses toward short-term rewards. However, note that ‘conflict’ applies to a much wider range of phenomena than this. Any creature with a heterogeneous set of goals that cannot be simultaneously pursued will need self-control. And the conflicts that get monitored might simply be conflicts between pursuing one or another goal at any particular time. Hence, there is a need for conflict monitoring even in the absence of being distracted or tempted.[[7]](#footnote-7)

One final objection to consider concerns empirical evidence that self-control and cognitive control are distinct constructs. In a recent study, Scherbaum et al. (2018) proposed that self-control and cognitive control recruit distinct processes and are, therefore, distinct constructs. They hypothesized that if cognitive control and self-control rely on the same set of processes, then “experimental manipulations that cause control adjustments in one task should also increase controlled behavior in the other task” (2018: 195). The experimental design used two tasks, a cognitive control task (Simon task) and a self-control task (intertemporal value decision). Each trial pair consisted of the cognitive control task followed by the self-control task. Participants had two seconds to perform the Simon task before immediately moving to the self-control task. Every participant completed 576 trial pairs. The results showed that there is “no difference in the probability of [choices exhibiting high self-control] for decisions following conflict Simon trials…and decisions following non-conflict Simon trials” (2018: 196). Hence, they conclude that self-control and cognitive control recruit distinct processes.

There are two problems with this, however. The first is a small point with experimental design. There is some evidence that when individuals perform a control-demanding task, they exhibit a tendency to continue performing the same task despite being cued to switch. This is known as the ‘task-set inertia’ effect (Allport et al., 1994; Yeung, 2010). The experimental design, however, did not rule out the possibility of task-set inertia effects. Hence, participants might have directed control resources toward the Simon task, relying on more habitual/default routines to complete the self-control task. The short time scale over which participants completed trials does not compensate for potential inertia effects.

The second problem concerns the hypothesis, namely that if cognitive control and self-control share similar processes, then tasks that activate cognitive control processes should elicit higher degrees of self-control in subsequent tasks. This, however, does not hold true even in cases where the two tasks are cognitive control tasks. For instance, Yeung et al. (2006) found that participants perform a cognitive control task more slowly and less accurately following the performance of a different cognitive control task when compared to performance of the same cognitive control task (there were also small ‘restart’ costs observed on same-same cognitive control task blocks). Hence, there is no reason to expect the claim that if two constructs share similar processes, then performing one kind of task will boost performance of the other kind of task. For these reasons, we can dismiss the supposed evidence in favor of the distinction between cognitive control and self-control.

With this model in hand, we can explain some phenomena connected with traditional studies of self-control. The first concerns the relationship between self-control and effort. The underlying neurological basis of self-control, on this view, is the cognitive control system implicated in maintaining task sets. Allocations of control, on this view, are determined by cost-benefit analyses and so are partly sensitive to the costs of control. This provides a naturalistically plausible framework for thinking about the feeling of effort. Effort is the subjective correlate or phenomenological component associated with the costs of control, specifically the opportunity costs associated with allocating control in a particular way (Cohen, 2017; Kool et al., 2017-b; Kurzban et al., 2013; Shenhav et al., 2017). However, this experience of effort need not always be aversive. This is because sometimes the lost opportunity costs are not valued more than the goals pursued in allocating control. When control is aversive, this might be due to diminished *motivation* or willingness to pay the costs of control (see Botvinick & Braver, 2015; Winecoff & Huettel, 2017). This would then incorporate a motivational element into the ‘self-control as vigilance theory’ (or, Vigilance Theory), thereby explaining the appeal of purely motivational views of self-control posited before.

Secondly, the view shows how impulse control/temptation resistance falls under the purview of self-control. Succumbing to temptation or giving into impulse is a kind of recovery failure. Typically, impulses and temptations have pull in virtue of being goal-relevant (though these goals might be maximally coarse-grained, like ‘experience pleasure’; see Levy, 2011: 143). Hence, being in the presence of a temptation or under the pull of an impulse threatens to shift goal pursuit. Self-control is needed to recover from the initial pull of temptations and impulses. This recovery consists in reallocating control to manage conflicts between the goals activated by temptations/impulses and the goals one was originally pursuing.[[8]](#footnote-8)

Ego depletion effects also make sense on the vigilance view. In general, the Vigilance Theory predicts that in a traditional two-task paradigm (like those used in original ego depletion studies) agents lose motivation to perform the task rather than run out of metabolic resources. In other words, the opportunity costs associated with allocating control to task performance becomes too high and agents shift either to focused thinking about something else or mind wandering (Murray et al., 2020). This connects to an old interpretation of the vigilance decrement (the deficit in performance associated with continued performance of a vigilance task), namely that as the stimuli associated with task performance loses novelty, the subject begins to think of other things (Broadbent, 1958). In other words, when agents get bored, they shift to other tasks and therefore perform worse.

This also answers some criticisms addressed to attention/motivation theories of self-control. For instance, Baumeister & Vohs (2016: 100) offer two criticisms:

[Some suggest] that fatigue in general has nothing to do with low energy but is instead a signal to interrupt one’s activities, reflecting opportunity and regulatory costs of perseverance. But what are those regulatory costs if not expenditure of energy?

…The opportunity cost argument has difficulty explaining the multi-task paradigm findings. If fatigue were merely a signal that it is generally a good idea to switch tasks (as opposed to being a signal that one’s energy has been somewhat depleted), why would it transfer so that fatigue from the first task is still felt during the unrelated second task (and indeed impairs performance on it?).

Consider these criticisms in reverse order. The second criticism we already considered by suggesting that subjects get bored and shift attention. The reason why this effect carries over between tasks is that subjects lose their motivation to perform in-lab activities (Nicholls et al., 2015). This also explains why increasing incentives on task performance between tasks in the ego depletion paradigm eliminates the depleting effect entirely (Muraven & Slessareva, 2003). When an agent sufficiently values performance on the second task, the opportunity costs reset and the agent performs close to ceiling (there are also reports of limited feelings of mental effort in these cases; see Boksem and Tops, 2008; Lorist et al., 2005).

The first criticism is simply a demand to understand what grounds the cost of control if not depletion of a metabolic resource. Recall from Section 2 that the ground of opportunity costs is the shared representational resources used in the brain (i.e., multiplexing). Because of the limited representational resources available for use in cognition, there are costs associated with allocating control (devoting representational resources to one processing pathway over another) that give rise to opportunity costs. Hence, the theory of self-control offered here can answer the criticisms raised against similar views.

*Conclusion*

This chapter showed three things. The first is that the standard story of self-control is incorrect. When we reorient our understanding of self-control around trying to understand limitations associated with control that derive from representational resources (rather than metabolic resources or brute structural limitations), a biologically plausible view of self-control emerges based on cost-benefit analyses of allocating control to a single task.

We also saw how this theory of self-control connects to vigilance. As vigilance manages goal maintenance, and self-control consists in allocating control to maintain goal pursuit, self-control is a species of vigilance (the two are not equivalent, as monitoring is not reducible to maintenance or self-control functions). There is a way to translate self-control functions into the language of vigilance when we characterize self-control functions in terms of maintenance. This also suggested a way to connect the sub-personal computational story about the allocation of control to personal-level properties. In particular, a crucial aspect of maintenance is task construal, and the representation of certain task sets at the computational level is sensitive to personal-level construal. Hence, vigilance (or the self-control aspect of vigilance) is not just a matter of subperonal computations.

Finally, we saw how the vigilance theory of self-control explains several phenomena related to self-control such as effort and ego depletion. Additionally, the vigilance theory can supply answers to criticisms normally leveled against motivation/attention views of attention.

Acting over time is different from concatenated sequences of actions at a time. This shows that many elements of planning psychology are temporally modulated (or have future-directed aspects). From this perspective, one issue with the standard story of self-control is that it centers on the needs of agents acting *at a time*. Even discussions of resolutions and temporally extended wills are all about preparing to act at a time. A planning agent needs something different. This account of self-control takes acting *in time* as fundamental to self-control. This is reflected in the EVC computation containing an essential future-oriented component. This is part of a mechanism for future-directed self-control and, hence, fits the needs of a planning agent.

**References**

Ainslie, G. and Haslam, N. 1992. “Self-control,” *Choice over Time*, Eds. G. Loewensten and J. Elster (New York: Russell Sage Foundation), 177-209.

Allport, A., Antonis, B., and Reynolds, P. 1972. “On the division of attention: A disproof of the single channel hypothesis,” *Quarterly Journal of Experimental Psychology* 24:2, 225-235.

Allport, A., Styles, E.A., and Hsieh, S. 1994. “Shifting intentional set: Exploring the dynamic control of tasks,” In C. Umiltà and M. Moscovitch (eds.), *Attention and performance XV: Conscious and nonconscious information processing* (Cambridge, MA: MIT Press), 421-452.

Anderson, M. 2014. *After Phrenology* (Cambridge, MA: MIT Press).

Baddeley, A. 2012. “Working Memory: Theories, Models, and Controversies,” *Annual Review of Psychology* 63, 1-29.

Bargh, J.A. and Chartrand, T.L. 1999. “The Unbearable Automaticity of Being,” *American Psychologist* 54:7, 462-79.

Baumeister, R.G., Bratslavsky, E., Muraven, M., and Tice, D.M. 1998. “Ego depletion: Is the active self a limited resource?” *Journal of Personality and Social Psychology* 74, 1252-1265.

Baumeister, R.F. and Vohs, K.D. 2016. “Strength Model of Self-Regulation as Limited Resource: Assessment, Controversies, Update,” *Advances in Experimental Social Psychology* 54, 67-127.

Bogacz, R., Brown, E.T., Moehlis, J., Hu, P., Holmes, P., and Cohen, J.D. 2006. “The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced choice tasks,” *Psychological Review* 113:4, 700-765.

Boksem, M.A.S. and Tops, M. 2008. “Mental fatigue: Costs and Benefits,” *Brain Research Reviews* 59:1, 125-39.

Botvinick, M.M. and Braver, T.S. 2015. “Motivation and cognitive control: from behavior to neural mechanism,” *Annual Review of Psychology* 3:66, 83-113.

Botvinick, M.M., Braver, T.S., Carter, C.S., Barch, D.M., and Cohen, J.D. 2001. “Conflict monitoring and cognitive control,” *Psychological Review* 108:3, 624-652.

Braem, S. 2017. “Conditioning task switching behavior,” *Cognition* 166, 272-76.

Broadbent, D.E. 1958. *Perception and communication* (Elmsford, NY: Pergamon Press).

Carter, E.C., Kofler, L.M., Forster, D.E., and McCullough, M.E. 2015. “A series of meta-analytic tests of the depletion effect: Self-control does not seem to rely on a limited resource,” *Journal of Experimental Psychology: General* 144:4, 768-815.

Carter, E.C. and McCullough, M.E. 2014. “Publication bias and the limited strength model of self-control: has the evidence for ego depletion been overestimated?” *Frontiers in Psychology* 5, 823.

Cohen, J.D. 2017. “Cognitive Control: Core Constructs and Current Considerations,” *Wiley Handbook of Cognitive Control*, Ed. T. Egner (London: Wiley), 3-28.

Craver, C. 2007. *Explaining the Brain* (Oxford: Oxford University Press).

Dang, J. 2018. “An updated meta-analysis of the ego depletion effect,” *Psychological Research* 82:4, 645-51.

Davisson, E.K. and Hoyle, R.H. 2017. “The Social-Psychological Perspective on Self-Regulation,” *Wiley Handbook of Cognitive Control*, Ed. T. Egner (London: Wiley), 440-454.

DeWall, C.N., Baumeister, R.F., Mead, N.L., and Vohs, K.D. 2011. “How leaders self-regulate their task performance: Evidence that power promotes diligence, depletion, and disdain,” *Journal of Personality and Social Psychology* 100, 47-65.

Dunn, T.L., Lutes, D.J.C., and Risko, E.F. 2016. “Metacognitive evaluation in the avoidance of demand,” *Journal of Experimental Psychology: Human Perception and Performance* 42:9, 1372-87.

Eastwood, J.D., Frischen, A., Fenske, M.J., and Smilek, D. 2012. “The unengaged mind: Defining Boredom in Terms of Attention,” *Perspectives in Psychological Science* 7:5, 482-95.

Edin, F., Klingberg, T., Johansson, P., McNab, F., Tegner, J., and Compte, A. 2009. “Mechanism for top-down control of working memory capacity,” *Proceedings of the National Academy of Sciences of the United States of America* 106:16, 6802-6807.

Etherton, J.L., Osborne, R., Stephenson, K., Grace, M., Jones, C., and De Nadai, A.S. 2018. “Bayesian analysis of multimethod ego-depletion studies favours the null hypothesis,” *British Journal of Social Psychology* 57:2, 367-385.

Evans, D., Boggero, I., and Sergerstrom, S. 2016. “The nature of self-regulatory fatigue and “ego depletion”: Lessons from physical fatigue,” *Personality and Social Psychology Review* 20:4, 291-310.

Feng, S.F., Schwemmer, M., Gershman, S.J., and Cohen, J.D. 2014. “Multitasking vs. multiplexing: Toward a normative account of limitations in the simultaneous execution of control-demanding behaviors,” *Cognitive, Affective and Behavioral Neuroscience* 14:1, 129-46.

Forbus, K.D., Gentner, D., and Law, K. 1995. MAC/FAC: A model of similarity-based retrieval,” *Cognitive Science* 19, 141-205.

Fujita, K. 2011. “On conceptualizing self-control as more than the effortful inhibition of impulses,” *Personality and Social Psychology Review* 15:4, 352-66.

Fujita, K., Trope, Y., Liberman, N., and Levin-Sagi, M. 2006. “Construal levels and self-control,” *Journal of Personality and Social Psychology* 90, 351-367.

Gailliot, M. and Baumeister, R.F. 2007. “The physiology of willpower: Linking blood glucose to self-control,” *Personality and Social Psychology Review* 11, 303-327.

Gailliot, M., Baumeister, R.F., DeWall, C.N., Maner, J.K., Plant, E.A., Tice, D.M., Brewer, L.E., and Schmeichel, B. 2007. “Self-control relies on glucose as a limited energy source: Willpower is more than a metaphor,” *Journal of Personality and Social Psychology* 92, 325-336.

Gershman, S.J., Horvitz, E.J., and Tenenbaum, T.B. 2015. “Computational rationality: A converging paradigm for intelligence in brains, minds, and machines,” *Science* 349:6245, 273-278.

Gollwitzer, P. 1999. “Implementation intentions: Strong effects of simple plans,” *American Psychologist* 54, 493-503.

Hagger, M.S., Chatzisarantis, N.L.D., Alberts, H., Anggono, C.O., Batailler, C., Birt, A.R., et al. 2016. “A Multilab Preregistered Replication of the Ego-Depletion Effect,” *Perspectives in Psychological Science* 11:4, 546-73.

Hagger, M.S., Wood, C., Stiff, C., and Chatzisarantis, N.L. 2010. “Ego depletion and the strength model of self-control: A meta-analysis,” *Psychological Bulletin* 136, 495-525.

Hinton, G.E., McClelland, J.L., and Rumelhart, D.E. 1986. *Parallel Distributed Processing. Volume I: Foundations* (Cambridge, MA: MIT Press).

Ho, M.K., Abel, D., Correa, C.G., Littman, M.L., Cohen, J.D., and Griffiths, T.L. 2022. People construct simplified mental representations to plan. *Nature* 606, 129-36.

Hockey, G.R.J. 2011. “A motivational control theory of cognitive fatigue,” *Cognitive fatigue: multidisciplinary perspectives on current research and future applications*, ed. P.L. Ackerman (Washington, D.C.: American Psychological Association), 167-88.

Holton, R. 2009. *Willing, Wanting, Waiting* (Oxford: Oxford University Press).

Inzlicht, M., Schmeichel, B.J., and Macrae, C.N. 2014. “Why self-control seems (but may not be) limited,” *Trends in Cognitive Sciences* 18, 127-33.

Inzlicht, M., Shenhav, A., Olivola, C.Y. 2018. “The Effort Paradox: Effort is Both Costly and Valued,” *Trends in Cognitive Sciences* 22:4, 337-349.

Joosten, A., van Dijke, M., Van Hiel, A., De Cremer, D. 2015. “Out of Control!? How Loss of Self-Control Influences Prosocial Behavior: The Role of Power and Moral Values,” *PLoS ONE* 10:5, e0126377.

Keough, K.A., Zimbardo, P.G., and Boyd, J.N. 1999. “Who’s Smoking, Drinking and Using Drugs? Time Perspective as a Predictor of Substance Use,” *Basic and Applied Social Psychology* 21, 149-64.

Kool, W. and Botvinick, M.M. 2014. “A labor/leisure trade-off in cognitive control,” *Journal of Experimental Psychology: General* 143, 131-141.

Kool, W., Gershman, S.J., and Cushman, F.A. 2017-a. “Cost-benefit arbitration between multiple reinforcement-learning systems,” *Psychological Science* 28, 1321-1333.

Kool, W., McGuire, J.T., Rosen, Z.B., and Botvinick, M.M. 2010. “Decision making and the avoidance of cognitive demand,” *Journal of Experimental Psychology* 139:4, 665-682.

Kool, W., Shenhav, A., and Botvinick, M.M. 2017-b. “Cognitive Control as Cost-Benefit Decision Making,” *Wiley Handbook of Cognitive Control*, Ed. T. Egner (London: Wiley), 167-89.

Kurzban, R. 2016. “The sense of effort,” *Current Opinion in Psychology* 7, 67-70.

Kurzban, R., Duckworth, A., Kable, J.W., and Myers, J. 2013. “An opportunity cost model of subjective effort and task performance,” *Behavioral and Brain Sciences* 36, 661-679.

LeCun, Y., Bengio, Y., and Hinton, G. 2015. “Deep learning,” *Nature* 521, 436-444.

Levy, N. 2011. “Resisting ‘Weakness of the Will’,” *Philosophy and Phenomenological Research* 82:1, 134-55.

Lieder, F., Shenhav, A., Musslick, S., Griffiths, T.L. 2018. “Rational metareasoning and the plasticity of cognitive control,” *PLoS Computational Biology* 14:4, 31006043.

Lorist, M.M., Boksem, M.A., Ridderinkhof, K.R. 2005. “Impaired cognitive control and reduced cingulate activity during mental fatigue,” *Cognitive Brain Research* 24:2, 199-205.

Ma, W.J. and Huang, W. 2009. “No capacity limit in attentional tracking: Evidence for probabilistic inference under a resource constraint,” *Journal of Vision* 9:11, 1-30.

Mele, A. 2011. “Self-Control in Action,” *The Oxford Handbook of the Self*, Ed. S. Gallagher (New York: Oxford University Press).

Miles, E., Sheeran, P., Baird, H., Macdonald, I., Webb, T., Harris, P. 2016. “Does self-control improve with practice? Evidence from a 6-week training program,” *Journal of Experimental Psychology: General* 145:8, 1075-1091.

Miller, E.K. and Cohen, J.D. 2001. “An integrative theory of prefrontal cortex function,” *Annual Review of Neuroscience* 24, 167-202.

Miller, G.A. 1956. “The magical number seven, plus or minus two: Some limits on our capacity for processing information,” *Psychological Review* 63, 81-97.

Mischel, W., Shoda, Y., and Rodriguez, M.I. 1989. “Delay of gratification in children,” *Science* 244, 933-938.

Muraven, M. 2003. “Prejudice as self-control failure,” *Journal of Applied Social Psychology* 38, 314-33.

Muraven, M. and Baumeister, R.F. 2000. “Self-regulation and depletion of limited resources: Does self-control resemble a muscle?” *Psychological Bulletin* 126, 247-259.

Muraven, M. and Slessareva, E. 2003. “Mechanisms of self-control failure: Motivation and limited resources,” *Personality and Social Psychology Bulletin* 29, 894-906.

Murray, S. 2017. “Responsibility and Vigilance,” *Philosophical Studies* 174:2, 507-527.

Murray, S. 2020. The place of the trace: Negligence and responsibility. *Review of Philosophy and Psychology* 11:1, 39-52.

Murray, S. and De Brigard, F. 2021. The neurocognitive mechanisms of responsibility: A framework for normatively relevant neuroscience. In M. Hevia (ed.) *Developments in Neuroethics and Bioethics*, vol. 4 (New York: Academic Press), 19-40.

Murray, S., Krasich, K., Schooler, J.W., and Seli, P. 2020. What’s in a task? Complications in the study of the task-unrelated-thought variety of mind wandering. *Perspectives on Psychological Science* 15:3, 572-88.

Murray, S. and Vargas, M. 2020. Vigilance and control. *Philosophical Studies* 177, 825-843.

Musslick, S. and Cohen, J.D. 2021. Rationalizing constraints on the capacity for cognitive control. *Trends in Cognitive Sciences* 25:9, 757-775.

Myrseth, K.O.R., and Fishbach, A. 2009. “Self-Control: A Function of Knowing When and How to Exercise Restraint,” *Current Directions in Psychological Science* 18:4, 247-52.

Nicholls, M.E.R., Loveless, K.M., Thomas, N.A., Loetscher, T., and Churches, O. 2015. “Some participants may be better than others: Sustained attention and motivation are higher early in semester,” *Quarterly Journal of Experimental Psychology* 68:1, 10-18.

Oberauer, K. and Kliegel, R. 2006. “A formal model of capacity limits in working memory,” *Journal of Memory and Language* 55, 601-26.

Pezzulo, G., Rigoli, F., and Chersi, F. 2013. “The Mixed Instrumental Controller: Using Value of Information to Combine Habitual Choice and Mental Simulation,” *Frontiers in Psychology* 4, 92.

Rougier, N.P., Noelle, D.C., Braver, T.S., Cohen, J.D., and O’Reilly, R.C. 2005. “Prefrontal cortex and the flexibility of cognitive control: Rules without symbols,” *Proceedings of the National Academy of the United States of America* 102:20, 7338-7343.

Sabb, F.W., Bearden, C.E., Glahn, D.C., Parker, D.S., Freimer, N., and Bilder, R.M. 2008. “A collaborative knowledge base for cognitive phenomics,” *Molecular Psychiatry* 13:4, 350-60.

Scherbaum, S., Frisch, S., Holfert, A.-M., O’Hora, D., and Dshemuchadse, M. 2018. “No evidence for common processes of cognitive control and self-control,” *Acta Psychologica* 182, 194-99.

Schmeichel, B. 2007. “Attention control, memory updating, and emotion regulation temporarily reduce the capacity for executive control,” *Journal of Experimental Psychology* 136, 241-255.

Schmeichel, B., Vohs, K.D., and Baumeister, R.F. 2003. “Intellectual performance and ego depletion: Role of the self in logical reasoning and other information processing, “*Journal of Personality and Social Psychology* 85, 33-46.

Schmeichel, B., Cohs, K.D., and Duke, S.C. 2011. “Self-control at high and low levels of mental construal,” *Social Psychological and Personality Science* 2:2, 182-189.

Shenhav, A., Botvinick, M.M., and Cohen, J.D. 2013. “The expected value of control: An integrative theory of anterior cingulate cortex function,” *Neuron* 79, 217-240.

Shenhav, A., Musslick, S., Lieder, F., Kool, W., Griffiths, T.L., Cohen, J.D., and Botvinick, M.M. 2017. “Toward a rational and mechanistic account of mental effort,” *Annual Reviews of Neuroscience* 40, 99-124.

Soutschek, A., Ruff, C.C., Strombach, T., Kalenscher, T., and Tobler, P.N. 2016. “Brain stimulation reveals crucial role of overcoming self-centeredness in self-control,” *Science Advances* 2:10, e1600992.

Sripada, C., Kessler, D., and Jonides, J. 2016. “Sifting Signal from Noise with Replication Science,” *Perspectives on Psychological Science* 11:4, 576-78.

Usher, M., Cohen, J.D., Haarmann, H., and Horn, D. 2001. Neural mechanism for the magical number 4: Competitive interactions and nonlinear oscillation. *Behavioral and Brain Sciences* 24:1, 151-52.

Vohs, K.D. and Faber, R. 2007. “Spent resources: Self-regulatory resource availability affects impulse buying,” *Journal of Consumer Research* 33, 537-47.

Watson, G. 1977. “Skepticism about Weakness of Will,” *Philosophical Review* 85, 316-39.

Wenzel, M., Lind, M., Rowland, Z., Zahn, D., and Kubiak, T. 2019. “The limits of ego depletion: A crossover study on the robustness of performance deterioration in consecutive tasks,” *Social Psychology* 50:5-6, 292-304.

Winecoff, A.A., and Huettel, S.A. 2017. “Cognitive Control and Neuroeconomics,” *Wiley Handbook of Cognitive Control*, Ed. T. Egner (London: Wiley), 408-421.

Yeung, N. (2010). “Bottom-up influences on voluntary task switching: The elusive homunculus escapes,” *Journal of Experimental Psychology: Learning, Memory, and Cognition* 36:2, 348-62.

Yeung, N., Nystrom, L.E., Aronson, J.A., and Cohen, J.D. 2006. Between-task competition and cognitive control in task switching. *Journal of Neuroscience* 26:5, 1429-38.

1. Spinoza captures this point eloquently: “How would it be possible, if salvation were ready to our hand, and could without great labour be found, that it should be by almost all men neglected? But all things excellent are as difficult as they are rare” (*Ethics* Vp52 s). [↑](#footnote-ref-1)
2. Feng et al. (2014) point out that self-control processes depend on prefrontal cortex. However, other cognitive and perceptual processes (like vision) engage the prefrontal cortex for sustained periods of time without any depletion effects on the scale observed in self-control studies. [↑](#footnote-ref-2)
3. Self-control is just a species of vigilance, not equivalent to it. This is because some exercises of vigilance might not consist in a manifestation of self-control. Again, this will be a point of discussion in Section 4. [↑](#footnote-ref-3)
4. Monitoring encompasses implementation operations because monitoring aims to bring about implementation in the right context. Therefore, there are two dimensions of vigilance rather than three. [↑](#footnote-ref-4)
5. There is some evidence that low-level construal sometimes benefits self-control (see Schmeichel et al., 2011). [↑](#footnote-ref-5)
6. See the discussion of constitutive relevance in Craver (2007). [↑](#footnote-ref-6)
7. This contradicts Gary Watson’s assessment of the role that the virtue of self-control plays within the moral life: “Self-control is a virtue only for beings who are susceptible to motivation which is in potential conflict with their judgments of what is good to pursue” (1977: 322). Self-control, as a component of vigilance, is a virtue only for planning agents that must manage pursuit of heterogeneous goods across time. [↑](#footnote-ref-7)
8. There is some question as to what distinguishes temptations and impulses from ordinary conflicts that arise from plural goal pursuit. This, I think, is not a question that falls on the Vigilance Theory to answer, but I suspect the answer has something to do with agential autonomy. That is, something is a temptation or impulse in virtue of the fact that the goal pursuit activated by the temptation or impulse is either counter-resolutional (as in Holton, 2009) *or* that the agent does not identify or endorse the goal pursuit (as in cases of addiction relapse). This, however, has more to do with the metaphysics of autonomy and the psychology of commitments than with self-control. [↑](#footnote-ref-8)