THE ATOMS OF SELF-CONTROL

Chandra Sripada

University of Michigan

Philosophers routinely invoke self-control in their theorizing, but major questions remain about what exactly self-control is. I propose a componential account in which an exercise of self-control is built out of something more fundamental: basic intrapsychic actions called cognitive control actions. Cognitive control regulates simple, brief states called response pulses that operate across diverse psychological systems (think of one's attention being grabbed by a salient object or one's mind being pulled to think about a certain topic). Self-control ostensibly seems quite different because it regulates complex, temporally extended states such as emotions and cravings. But critically, these complex states also exhibit important componential structure: They rely on response pulses as a key means by which they bring about action. The overall picture is that self-control consists of skilled sequences of cognitive control directed against extended streams of response pulses that arise from states such as emotions and cravings, thus preventing these states from being effective in action. The account clarifies the "atoms" of self-control—the elemental units that get combined in complex ways to produce different kinds of self-control actions. Surprisingly, the account, which is derived from research in cognitive science, aligns nicely with the commonsense conception of self-control.

1. Introduction

Even when famished, a person can resist their desire to eat. Though an approaching pit bull is absolutely terrifying, a person can stop themselves from fleeing. In the midst of a truly boring lecture, a person can focus on the professor and resist the ongoing temptation to peek at their phone.

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/nous.12332.

These are all paradigm examples of *self-control*, in particular, synchronic self-control, self-control directed at a desire that is currently active.[1] Self-control is widely recognized in philosophy as a central agential capacity, and it is routinely invoked in theorizing about free will, moral responsibility, weakness of will, and diachronic rationality, among other topics. Yet philosophical accounts aimed at systematically explicating the capacity for self-control remain relatively rare. Major questions remain about what self-control is and how it works.

My aim in this paper is to build a unified, mechanistically precise account of self-control. In undertaking this task, I draw heavily on a major research program that has emerged in neuroscience and psychology that studies the phenomenon of *cognitive control*. The core of this research is systematic investigation into mechanisms that underlie performance in "conflict tasks", a large set of tasks that involve regulating a variety of spontaneous mental phenomena, including actions that arise habitually, attention that is grabbed by stimuli, memory items that are automatically retrieved, and thought contents that spontaneously pop into mind.

Research into cognitive control, with its emphasis on simple responses in carefully constructed laboratory tasks, tends to be somewhat far removed from questions about self-control as it operates in the real world to regulate complex, extended motivational states such as emotions and cravings. While it is widely thought that the two have *something* in common, the connections have not yet been made explicit and systematic.[2]

Thus a major task of the paper is to build the needed bridge between cognitive control and self-control. With the bridge in place, the result is a unified, empirically grounded account of self-control, what I call the "atomic" model. The key idea is that cognitive control and self-control are related componentially: exercises of self-control directed at complex motivational states, such as emotions and cravings, consist in executing extended, skilled sequences of cognitive control. The account illuminates the atoms of self-control—the elemental units that are assembled together in complex and diverse ways to produce self-control actions. By taking a componential perspective, the account helps us see why even though exercises of self-control take diverse forms and are directed at diverse targets, they are all still members of a unified theoretical kind.

---

[1] This contrasts with what some call "diachronic self-control", which prevents an unwanted, non-occurrent, anticipated future desire from becoming active. Going forward, when I use the term "self-control" without a qualifier, I am referring exclusively to synchronic self-control. In §6.3, I clarify the difference between synchronic and diachronic self-control, and I suggest that pure diachronic self-control is a misnomer—it isn't a form of self-control at all.

[2] Attempts to bridge certain cognitive control-related constructs with self-control are found in Robinson, Schmeichel and Inzlicht, 2010 and Hofmann, Schmeichel and Baddeley, 2012.

The remainder of the paper is divided into five parts. Part 2 sets out a larger context for the problem of self-control and homes in on the specific issue I aim to address. Because the account I will to set out in this paper has several relatively separate pieces, I also provide an overview of the overall account there. Part 3 presents theory and findings from the cognitive control research program. Part 4 offers a theory of emotions, drives, cravings, and related states, which in turn sets up a bridge between cognitive control and self-control. Part 5 formally sets out the atomic model of self-control. Part 6 examines the atomic model in light of the commonsense conception of self-control, and part 7 concludes.

2. Setting the Stage and an Overview of What Is to Come

2.1. The Charioteer Metaphor

In a classic historical formulation of the problem of self-control, the *Ratha Kalpana*, the third chapter of the *Katha Upanishad,* puts forward a metaphor in which a charioteer, the intellect, can follow the correct path only by "subduing" ill-behaved horses, the senses (Deussen, 1980). Plato presents a similar metaphor in the *Phaedrus* in which a charioteer drives two horses. One horse is of lowly breed and so "the driving is necessarily difficult and troublesome" (Plato, 1952, secs. 256b, 253d–254).

The charioteer metaphor provides a helpful overall context for the problem of self-control, which is usefully conceptualized in terms of three elements. First, there are the states and processes that make up the controller, i.e., the charioteer. Second, there are the states and processes that are the things that are controlled, i.e., the horses. The third element consists of the states and processes that are the means by which the controller regulates the things that are controlled. In the metaphor, this is the role played by the whip or the reins, and it is what we can call the *regulative* aspect of the problem of self-control.

Over the ensuing centuries, there has been much philosophical work relating to the first two elements. For example, there has been much attention directed at characterizing faculties of reasoning and deliberation, which are closely connected to the controller. There has also been extensive work elucidating the states that are targets of control, for example desires, emotions,

and pleasures/pains. There is notably much less written, however, about the regulative aspect of self-control.[3]

This is not to say this third element is unimportant for philosophical theorizing. Theorists discussing a variety of topics in moral psychology—moral responsibility, weakness of will, intertemporal choice—routinely appeal to some capacity we have to regulate occurrent wayward motivation; they invoke this idea frequently, sometimes in crucial places in their arguments.[4] But invoking a notion and explicating it are, of course, two very different things. My aim in this essay is to make some progress in understanding the regulative aspect of self-control, drawing on a sizable body of research in psychology and neuroscience on cognitive control that hasn't yet made much impact in philosophy.

2.2. A Roadmap

A distinctive feature of the account of self-control I put forward is that it asks us to take up a different perspective on the phenomenon, one that is much more "zoomed in" than usual. Consider exercising self-control over some motivational state, say a strong craving for a salty snack. It is more usual to see this phenomenon discussed from what we might call the "meso-scale", in which the exercise itself and the things that are controlled are seen as relatively basic entities. We don't typically shift to the "micro-scale" and focus in on the elemental units that compose these entities. The atomic model that I put forward, however, says that the goings on at this lower level hold the key to deep understanding of self-control.

My exposition of the atomic model unfolds in three main pieces. Each piece takes some time to set out, and it is only much later that they get assembled into an overall theory of self-control.

---

[3] Bibliometrics confirms this qualitative impression. Searches on Philosopher's Index show there is generally 10 to 50 times more work on topics such as "reasoning", "desire", "emotion", and "action" than "self-control". The area is, of course, not entirely neglected. Theorists in philosophy whose recent work has engaged with the regulative aspect of self-control include: Holton (2014; 2009; 2013), Kennett (2003; 1996), Levy (2011, 2013), Henden (2008; 2013), and Mele (1987, 1997, 2013).

[4] For example, consider appeals to "irresistible desires", i.e., desires that cannot be regulated no matter how hard the person tries, found routinely in the literature on moral responsibility. Or consider Gary Watson's famous claim that weakness of will is to be understood normatively—the weak-willed agent lacks the capacities to regulate wayward motivation that we should expect from a typical person (Watson, 1977). Or consider Jay Wallace's refutation of what he calls the "hydraulic" conception of agency by drawing attention to the ways we can control our occurrent desires, which the hydraulic view, it is claimed, rules out (Wallace, 1999).

For this reason, it will be helpful for the reader to receive a brief roadmap of what pieces will be coming and how they will eventually come together.

The first piece of my account of self-control, which comes in §3, is an account of *response pulses*, a notion that plays an absolutely central role in the cognitive control research program, but which thus far has never been given a detailed treatment. Response pulses are relatively simple and brief psychological states that dispose one to produce certain relatively simple responses—think of one's attention being grabbed by a salient stimulus or one's thoughts being pulled to a certain theme. I give a detailed account of response pulses as they arise in mechanisms associated with attention, memory, thought, and action.

The second piece of my account, also found in §3, is a discussion of a control actions. These are rapidly executed mental actions that target response pulses and prevent their associated response from occurring. I describe three major kinds of control actions, and fill in a broader picture of the "executive" mechanisms that organize how they are selected and executed.

The third piece of my account, which comes in §4, is an account of states such as emotions, urges, cravings, and other similar states—what I call "emotion-type states". These are temporally-extended states that are much more complex than response pulses, but they are nonetheless intimately connected with them. In particular, emotion-type states are potent sources of temporally-extended steams of activated response pulses associated with attention, memory, thought, and action. These response pulses, in turn, are a key means by which emotion-type states influence action.

With these pieces in place, the microscale perspective on self-control that I have in mind comes into view. The atomic model says that to understand what is going in an exercise of self-control directed at an emotion-type state, we must focus in on the substructure of both of these things—that is, we must examine the micro-level goings on of both the exercise of self-control as well as the emotion-type state that is targeted. The exercise of self-control, it is argued, consists of skilled sequences of cognitive control actions. The emotion-type state targeted is a source of temporally-extended streams of activated response pulses, which is a major way it produces action. Blocking these response pulses is how the exercise of self-control prevents the emotion-type state from being effective in action. Indeed, I defend a stronger claim: The *only* states that are ever *directly* targeted during exercises of self-control are response pulses.

The final part of the paper zooms back out to the "meso-scale". I take up the question of how the empirically-based account of self-control that I put forward lines up with the commonsense

conception of self-control. The answer, somewhat surprisingly, is: quite well. The phenomenology of effort, I argue, helps to explain this tight correspondence.

3. The Cognitive Control Research Program

3.1. Conflict Tasks and Response Pulses

The core of the cognitive control research program consists of systematic psychological and neural investigation into a large number of conflict tasks, tasks that produce a characteristic kind of divergence between subjects' spontaneous responses to task stimuli and the appropriate responses given the task instructions. Here I summarize five such tasks, each of which features a different kind of spontaneous state that is the target of control.

Stroop Task (Stroop, 1935) – On each trial, subjects are shown a color word ("red", "blue") which is itself printed in an ink color. Subjects are asked to state the ink color of the word on all trials. On congruent trials, the color word and ink color match and it is relatively easy to get the right answer. On incongruent trials, the color word and ink color are discrepant, and subjects must exert control over their spontaneous tendency to read the word in order to select the correct response.

Go/No Go Task (Donders, 1969) – On each trial, subjects see a letter on the screen. Subjects are asked to press a button only if the letter is not "X" and withhold the button press if it an "X". Most of the letters are not "X", for example 90% not "X" to 10% "X". This skewed ratio leads to the development of a habit for button pressing. On trials where the stimulus is not "X", the button pressing habit facilitates correct responding. On "X" trials, subjects must be suppress this habit.

Anti-Saccade Task (Munoz & Everling, 2004) – On each trial, subjects are shown a cross on the screen that moves either left or right. On congruent trials, subjects are asked to look in the same direction as the movement. On incongruent trials, they are asked to look in the opposite direction of the movement. This requires suppressing the spontaneous tendency to shift one's gaze in the direction of a moving object.

Visual Distractor Task[5] (Melloni, Leeuwen, Alink, & Müller, 2012) – On each trial, subjects must find the uniquely oriented grating on an array of textured objects. In addition, one of the objects in the display is salient and spontaneously "pops out" because of its unique color.

[5] The authors did not give this task a canonical name, and so I suggest this name.

On congruent trials the unique grating coincides with the salient object. On incongruent trials, the unique grating does not coincide with the salient object, and subjects must overcome their spontaneous tendency to look at the salient object in order to produce the correct response.

Think/No Think Task (Anderson & Green, 2001) – During a practice session, subjects are trained to recall pairs of words (e.g., ROACH – ORDEAL; GUM – TRAIN). In the test session, they are given the first member of the pair. They are told that if the word appears in green ink, they are to think about the paired word. If the word appears in red ink, they must not think about the paired word. This requires that they suppress the spontaneous tendency to recall the associated word.

3.2. Response Pulses

All five of the preceding tasks involve the elicitation of what I will call *response pulses*. For example, in the Anti-Saccade task, there is a response pulse to shift one's gaze in the direction of the moving cross. Importantly, response pulses need not be directed only at overt bodily responses. In the Think/No Think task, for example, the relevant response pulse is directed at recalling the associated word, an internal "response" of memory retrieval systems. The notion of a response pulse is fundamental to understanding the cognitive control research program. Remarkably, however, a detailed account of this notion is not available, so I want to spend some time now providing such an account.

Importantly, response pulses will play several key roles in my account of self-control. Later, I will be arguing that response pulses are the means by which states such as emotions and cravings influence attention, belief, thought, and action, thus posing self-control problems. Also, later I will argue that response pulses are the proximal targets of all self-control actions—*all* self-control proceeds by means of controlling response pulses. Thus, clarifying the nature of response pulses is critical to understanding my positive proposal.

First, a response pulse is a state that arises—or as it is sometimes put, is "activated"—in a psychological mechanism in a certain stimulus context, and the state in turn helps to explain why the mechanism produces a certain response in that context. The psychological functional role of a response pulse is broadly akin to what philosophers call an "action-desire", a desire to perform some action straightaway (Mele, 2003). Consistent with a broadly causalist picture of action (Davidson, 1963), a response pulse is similarly poised to cause and sustain a certain response. Thus to the questions, why do people (usually) shift their gaze towards a moving object?, and why do they (usually) attend to the salient object in a visual display?, and so on for the other conflict tasks, the answer is: because there are response pulses to do these things that arise in the respective task-associated stimulus conditions.

Second, response pulses are extremely brief and simple states. The time between their initial activation and the execution of their associated responses is typically on the order of hundreds of milliseconds. This contrasts with motivational states standardly discussed in the philosophical literature—emotions, cravings, etc.—that typically last minutes to hours or longer. Emotions, cravings and similar states are also much more complex in that they produce broad coordinated alterations across diverse psychological systems (attention, memory, evaluation, action selection, etc.), which I will spend some time describing in §4. Response pulses, in contrast, are much simpler in that they operate on individual psychological mechanisms, as illustrated in each of the conflict tasks described above.

Third, the activation of a response pulse occurs irrespective of one's explicit goals or judgments. This idea is illustrated vividly by observing what happens after a person has completed a large number of trials of any one of the preceding tasks. The standard observation is that even though subjects are perfectly clear on the task instructions and are highly motivated to follow these instructions, response pulses for the respective inappropriate responses arise in each task trial after trial. So, for example, in incongruent trials of the Anti-Saccade task, even after hundreds of such trials, people still have a response pulse to shift their gaze towards the moving object, and this response pulse must be overridden each time.

Fourth, there is a particularly tight tie between a response pulse and the response it plays a role in bringing about: the occurrence of the response pulse-favored response is a *default* state for the relevant mechanism. That is, suppose in some psychological mechanism *M* placed in situation *S*, a response pulse to *R* is activated. Then, under normal conditions, *M* will *R* unless something intervenes to prevent this. *R* is thus the default response of *M* in *S*.

The idea of a default response is linked to a broader picture in which the operation of a broad array of psychological mechanisms is characterized by dividing the explanation into two parts. First there is an account of what happens if the mechanism is placed in some situation *S* and is left to itself, and second, there is an account of what happens if an "exogenous" factor intervenes.[6] A response pulse is a state of a mechanism that plays a crucial role in the first part of the explanation: it helps to explain what the mechanism will do in *S* when left to itself.

[6] This "default-interventionist" picture has a long history in cognitive science and is present in early formulations of cognitive control (e.g., Miller and Cohen, 2001). It is also important in dual-process theories of reasoning (e.g., Evans, 2007; Evans and Stanovich, 2013).

Fifth, a response pulse is a "prepotent" state. That is, the occurrence of an activated response pulse only leads to the associated response if nothing else intervenes. But something *can* intervene; the system is configured with a set of control actions—to be discussed presently—that can stop a response pulse from producing its associated response. The prepotent nature of response pulses marks a major asymmetry in our psychology. A bit later, I will discuss executive mechanisms, which are the source of control actions that target response pulses. The response-producing states in executive mechanisms are *not* prepotent—the activation of these states will, under normal conditions, lead to their respective responses because there aren't still higher-level structures that can intervene. A hallmark of response pulses, then, is their susceptibility to interventionist control.

Finally, response pulses have a distinctive phenomenology, which shows up most clearly under conditions of conflict in which the person has a standing goal that a response pulse works against. Consider, for example, the Visual Distractor task where the person has the goal of focusing on the orientation of the gratings. The response pulse elicited by the distracting salient object produces a subjective experience of being "pulled away"; one's attention feels "grabbed" by the salient object. A broadly similar "pulled" or "grabbed" phenomenology arises as well in the other conflict tasks described above.

3.3. Control Actions

Having given an account of what response pulses are, I now want to turn to characterizing how people deploy what I will call "control actions" to volitionally regulate them. Control actions are basic mental actions[7] that target an activated response pulse and prevent the its associated response from occurring[8]. Setting out an inventory of control actions is important for my project because later I will argue that control actions that target response pulses are the elemental units that compose exercises of self-control. That is, the following list of cognitive control actions is, on my view, also a list of the basic actions of self-control.

---

[7] They are at least relatively basic. Future research may reveal that they too involve the execution of still more fundamental actions.

[8] There are two ambiguities in my formulation of what control actions do that are deliberate, as I want to avoid taking sides on tricky issues that the current state of the science cannot resolve. First, there is a question of whether control actions suppress the activity of a problematic response pulse or they enhance the activity of a competing response pulse, or both (see Aron, Robbins and Poldrack, 2014 for a discussion of this point). Second, there is a question of whether the response pulse itself is suppressed or whether the characteristic response associated with the response pulse is blocked while the response pulse itself, qua prepotent state, is not affected. For my purposes, what matters is that a control action prevents a response pulse from issuing its characteristic response. I leave the answers to these two questions, which concern the mechanisms by which this happens, open.

While the taxonomy of control actions is a topic of ongoing investigation, we can, broadly speaking, distinguish three main families.

Motor Inhibition Actions – Many tasks, such as the Go/No Go task discussed above, probe the ability to inhibit an initially activated motor response pulse. A central finding from systematic investigation of such tasks is that stopping is not simply the cessation of going, but rather involves intentionally activating an independent "stopping" process (Aron, Robbins, & Poldrack, 2004, 2014). That is, to stop an initially activated motor response pulse, the person must intentionally engage a separate set of inhibition processes that suppress the response pulse.

Attention Actions – Attention is sometimes drawn spontaneously to salient objects in the environment: a baby's cry, a lascivious photo flashed on the corner of the screen, a single red dot in group of grey ones. This is "bottom-up" attention. Attention can also be allocated intentionally based on one's current goals, which is called "top-down" attention. These two kinds of attention are produced by largely distinct brain regions (Corbetta & Shulman, 2002) that implement distinct computations (Itti & Borji, 2015).

A person can deploy top-down attention to override bottom-up attention. For example, in tasks like the Visual Distractor Task, a person can volitionally direct top-down attention to a stimulus location even when bottom-up attention produces an initial spontaneous tendency that favors attending to a salient stimulus (EinhÃ, Rutishauser, & Koch, 2008). As another example, in the Stroop task, a person can direct attention to task-appropriate representations of stimulus-response mappings (what are called "task sets"), even when an inappropriate task set, i.e., the word reading task set, is initially spontaneously activated (Egner & Hirsch, 2005).

Memory/Thought Actions – People can perform actions that suppress or otherwise alter the accessibility of memories or thoughts (Anderson & Green, 2001; Anderson et al., 2004; Nee, Jonides, & Berman, 2007). The most studied mechanism is memory suppression, illustrated in the Think/No Think task (Anderson & Green, 2001). Like motor inhibition, memory suppression is not simply omitting a memory action, such as omitting to retrieve an item from memory. Rather, there is evidence that memory suppression involves activating a mechanism that directly suppresses activity in brain regions that are the basis of accessing stored memories (e.g., hippocampus), as well as

related regions in sensory cortex that are components of the memory trace itself (Gagnepain, Hulbert, & Anderson, 2017).[9]

Memory suppression is likely to be closely related to suppression of spontaneously arising thoughts. There is good evidence that the construction of the spontaneous stream of thought, sometimes called mind wandering or daydreaming, involves repeated cycles of memory retrieval operations, especially retrieval of material from episodic memory (Andrews-Hanna, Smallwood, & Spreng, 2014; Christoff, Irving, Fox, Spreng, & Andrews-Hanna, 2016). If this is right, to suppress spontaneously arising thought, a person will need to suppress the memory retrieval operations that produce this form of thought.[10]

3.4. Properties of the "Controller"

Thus far, I have described control actions, but I haven't said anything about how they are "intelligently" selected in order to regulate response pulses. I refer to the cluster of mechanisms that accomplish this function as executive mechanisms, and turn now to filling in some of their major features. To be clear, response pulses and control actions are the key ingredients that will feature in my account of self-control. I am nonetheless spending some time discussing executive mechanisms for two reasons. The first is that I want to tie cognitive control to intentional action, as opposed to its simply consisting of "operations" that passively unfold in one's psychology. For that, I need to link cognitive control to value calculation and decision, which I do below. A second goal is to strive for a bit of completeness. I want to fill in enough about executive mechanisms so that the overall architecture "makes sense", and the reader is not left with the feeling that there is a major homunculus, the controller, that has not been, and maybe cannot ever be, adequately discharged.

3.4.1. Working Memory and Information Asymmetries

[9] In some cases, an activated response pulse proceeds to being an actual response that is rapidly overridden. This appears to be the case in the Think/No Think task. Similar to the points I made in note 8, I allow that control actions might work either by preventing a response pulse's characteristic response from occurring or by rapidly cancelling a response once it occurs. For our purposes, nothing of importance is lost in allowing this and the exposition is made much more succinct.

[10] The suppression of thought has also been studied extensively using Daniel Wegner's classic "white bear" paradigm, in which subjects are asked to avoid thinking about a white bear (see Wenzlaff and Wegner, 2000 for a review).

Working memory is the capacity to store and manipulate information that isn't perceptually present (Baddeley, 2012; Baddeley & Hitch, 1974). The standard view among cognitive control theorists is that there is a characteristic information asymmetry between mechanisms that are the source of control actions (i.e., executive mechanisms) and the mechanisms that are the source of response pulses: the former have access to working memory while the latter do not (Miller & Cohen, 2001).[11]

This information asymmetry is the ultimate source of divergent response tendencies that are the hallmark of conflict tasks. These tasks present an unusual situation in which the experimenter gives subjects a highly novel set of instructions. Successful task performance requires that these instructions and other goals relevant to the task situation need to be held in working memory and brought to bear on task performance, which is something executive mechanisms can accomplish. Mechanisms that are the source of response pulses lack (direct) access to this information, which is why they continue to reliably generate inappropriate response tendencies trial after trial, in turn generating the need for regulation.

3.4.2. Monitoring

It is widely held that executive mechanisms are not continuously "online" but become active only intermittently when cognitive control is required (Botvinick, Braver, Barch, Carter, & Cohen, 2001).[12] For this type of set-up to work, there needs to be a mechanism in place that plays a *monitoring* role: it detects when response pulses being generated by other processes are in conflict with one's goals, including contextual goals held in working memory (e.g., the goal to state the ink color of the word in the Stroop task). There is substantial evidence from neuroimaging, animal work, and lesion studies that the mind houses a mechanism, likely located in the anterior cingulate cortex, that plays this monitoring role (Botvinick, Cohen, & Carter, 2004).

---

[11] I am referring here to *direct* access. It is possible to deploy strategies in which the contents of working memory are "relayed" to disparate mechanisms that lack direct access to working memory (Carruthers, 2015). For example, this happens when one uses goals and other information in working memory to undertake a detailed visualization of an upcoming threat, thus allowing spontaneous inference systems to rapidly identify causes and consequences of the threat. Strategies such as these create numerous pathways of indirect access to working memory, and I leave these complications aside for the present discussion.

[12] This formulation is typical in the literature, though I think it is not quite accurate. In my view, executive mechanisms operate during essentially all cognitive tasks, but there are sharp differences across simple versus complex tasks in *how much* executive processing is required. For the purposes of this essay, I leave this complication aside.

3.4.3. Expected Value of Control and Executive Decisions

Monitoring addresses the *when* question: When should executive mechanisms come online and potentially exercise control? There is still the *which* and *how* questions: Out of all the candidate control actions that might be performed, which ones should be performed, in what order and arrangement, and with what intensity? Addressing this question constitutes what has been called "the control specification problem".

There is a growing consensus that *subpersonal* cost/benefit calculation plays a central role in addressing the control specification problem (Kool, Shenhav, & Botvinick, 2017; Shenhav, Botvinick, & Cohen, 2013). The basic idea is that there is a set of cognitive routines that continuously estimate a quantity dubbed "expected value of control" (EVC) (Shenhav et al., 2013; Shenhav, Cohen, & Botvinick, 2016). EVC represents the total benefits from exercising control with respect to one's goals, including local, contextual goals held in working memory, versus the total costs of exercising control. Calculation of EVC is linked with feedback mechanisms that register the success of previous rounds of control in type-similar situations and modify the value assigned to candidate control actions accordingly.[13]

How do representations such as EVC, as well as potentially other kinds of representations, get together to lead to intentional exercises of control directed at altering one's response pulses? The answer is by means of decisions.[14] I understand forming a decision as a process of evidence accumulation. When evidence for performing an action accumulates sufficiently so that it reaches a certain critical threshold, for example when estimates of overall benefits exceed costs, or exceed them by a certain margin, a person-level action ensues: the person *makes a decision* to

[13] It bears emphasis that EVC calculations are performed via subpersonal routines. The idea is not that the person consciously and intentionally sets out to figure out the expected value of control. Rather, the relevant calculations occur non-deliberatively. Now, this does not imply that one's conscious judgments, for example, one's practical judgments that regulating an emotion or craving is the thing to do, are irrelevant to whether one exercises control. Rather, in a rational person, these judgments will likely be an important informational input to the processes that tabulate the benefits and costs of control. The relation between practical judgment and decisions to exercise control is a complex topic that I intend to take up in due course, but I cannot expand on it here.

[14] It bears emphasis that I am using decision here in a more minimal sense that follows the usage that prevails in computational neuroscience. The standard view in this field is that *every* action is preceded by a decision in this minimal sense. In the ordinary understanding, in contrast, decisions are seen as rarer, more deliberate, and based on consciously accessible reasons.

perform the action.[15] I refer to decisions to exercise control directed at altering one's response pulses as "executive decisions" to distinguish them from decision-type processes that occur in various other psychological systems.[16]

3.5. Summing Up

Here, then, is the overall picture. Response pulses are response-disposing states that arise across diverse psychological mechanisms, including mechanisms associated with action, attention, memory, and thought. Importantly, they arise irrespective of one's explicit goals and judgments and are the basis for default responses of the relevant mechanisms.

Our minds, however, are in addition equipped with a set of executive mechanisms linked to working memory. In situations where response pulses are inappropriate and conflict with one's overall goals, we can engage in cognitive control. In particular, based on EVC representations, as well as potentially other sources of information, we can make executive decisions to perform control actions, including inhibitional, attentional, and memory/thought actions. These control actions can prevent an initially activated response pulse from producing its associated response, and instead allow an alternative response to prevail.

The preceding overall picture is what I take to be a fairly standard view among theorists working on cognitive control (for reviews, see Miller and Cohen, 2001; Botvinick and Cohen, 2014; Cohen, 2017). The picture enjoys convergent support from a very broad set of methods including: behavioral studies, computational models of accuracy and reaction times, computational simulation methods, neuroimaging including task-based fMRI and resting state fMRI, animal studies, lesion studies, methods involving manipulations (for example, working memory load, pharmacological challenge, sleep deprivation), and individual-difference methods. This is the picture I will be assuming going forward.

4. The Targets of Self-Control

[15] This picture draws heavily from sequential sampling models, a now standard model of decision in cognitive science (for a review, see Forstmann, Ratcliff and Wagenmakers, 2016).

[16] Suppose a person's strongest desire is to phi. Can they still make an executive decision to perform control actions to stop themselves from phi-ing? I believe the answer is yes. This kind of divergence is possible because there is an important degree of motivational segregation between executive mechanisms that produce control actions and the motivational states that are the targets of regulation. A picture broadly along these lines is defended in Sripada, 2014.

At least on the surface, it is not obvious that cognitive control, as it operates in conflict tasks such as the Stroop task, has much to do with the exercises of self-control in everyday life. Cognitive control targets response pulses—simple, brief, "cool" states. These seem quite different from the states targeted by self-control, for example strong emotions or persistent cravings that last minutes or hours. This difference is presumably the reason why it seems fairly odd to say that a person performing the Stroop task "exercises self-control" against their tendency to read the word response.

I want to argue, nonetheless, that appearances here are deceiving. In what follows, I use the term "emotion-type states" to refer to the states that are the targets of self-control. I then show that emotion-type states are potent sources of response pulses associated with attention, belief, memory, thought, and action, and these response pulses are a major way that emotion-type states affect action. Demonstrating this link between emotion-type states and response pulses is important for my positive proposal because it sets up a bridge between cognitive control and self-control. I will eventually argue (in §5) that self-control stops emotion-type states from being effective in action by stopping their associated activated response pulses. My narrower goal for this part of the paper is to link emotion-type states to response pulses, and to link these response pulses to action.

4.1. The Targets of Self-Control Are Emotion-Type States

The characteristic targets of self-control are a diverse collection of states that includes emotions, drives, impulses, cravings, pains, itches, and feelings of fatigue, among others. Because an adequate term that encompasses this collection is not available[17], and because emotions are the paradigmatic members of the collection, I refer to the collection as "emotion-type states".

What unites this class of states? I propose that one important unifying feature is that they share a common "core architecture". This architecture includes characteristic ways the relevant states are elicited and characteristic consequences that ensue after elicitation, which I set out in Figure 1. I then expand on some key features of this architecture in the following sections. I begin by focusing exclusively on emotions, and then broaden my discussion to include other emotion-type states.

[17] A notable attempt at a neologism to address this very problem is Lowenstein (1996), which refers to this collection as "visceral factors".

<place figure 1 here>

4.2. Emotions: Core Architecture and Biases Across Multiple Mechanisms

Start with the elicitation of emotions. According to contemporary theory, this involves *appraisals*, spontaneous interpretations of the situation with respect to one's standing basic concerns (Ellsworth & Scherer, 2003; Scherer, 2001). Appraisals are non-deliberative. For example, a person may explicitly believe, based on statistical evidence, that flying on planes is safe. But if they (non-deliberatively) appraise being on a plane as a threat, fear will be elicited. Additionally, appraisals are not just one-shot affairs; they are ongoing. Even after emotion elicitation, a person continues to spontaneously appraise the current situation, which in turn, depending on the contents of these appraisals, sustains, magnifies, or dampens the emotion episode.

The elicitation of an emotion state leads to widespread downstream consequences (Levenson, 1994). The most studied consequences include changes to physiological variables (e.g., elevated heart rate) and production of facial expressions, but these effects of emotions are not my focus in this article. I want to instead focus here on the effects of emotions on general features of cognition.

It is widely accepted that a key feature of emotions is that they produce characteristic *biases* on a number of psychological mechanisms. I summarize the major targets of emotion-produced biases below. My eventual goal here is to show that the biases produced by emotions are potent sources of temporally-extended streams of emotion-congruent response pulses associated with action/goal selection, attention, beliefs, memory, and thought.

Action/Goal Selection Mechanisms – A hallmark of emotions is they produce what theorists widely call "action tendencies" (Frijda, 1986). As we deal with day-to-day situational challenges, action/goal selection systems retrieve schemata for how to deal with these challenges. That is, these systems retrieve (fast and non-deliberatively) a best match schema containing higher-level goal structures and lower-level actions for how to respond. Emotions produce biases on schema retrieval, which in turn produce characteristic emotion-specific action tendencies—these biases favor action/goal

schema that, over evolutionary time, typically address the situational challenge that elicited the emotion (Frijda, 1986; Scarantino, 2014). For example, during fear, schema retrieval is strongly biased towards escape. During anger, it is biased towards retaliation. During sadness, it is biased towards submission and withdrawal, and so on for other emotions (Frijda, 1987; Frijda, Kuipers, & Ter Schure, 1989).

Attention Mechanisms –Bottom-up attention refers to attention that is spontaneously drawn to objects in the environment. This form of attention operates through the construction of salience maps, topographically arranged maps that assign objects in the environment a salience score. Emotions produce characteristic alterations in salience maps, creating biases of attention towards emotion-congruent stimuli. For example, due to attentional biases, under conditions of fear, threat-related stimuli are noticed more rapidly, reach consciousness more easily, and are subsequently better recalled (Öhman, Flykt, & Esteves, 2001; Phelps, Ling, & Carrasco, 2006).

Belief Formation/Evaluation Mechanisms – Most beliefs are formed spontaneously, without the need for explicit inference or deliberation (Uleman, Adil Saribay, & Gonzalez, 2008; McKoon & Ratcliff, 1992; Swinney & Osterhout, 1990). For example, you hear a noise downstairs, and it spontaneously occurs to you that your spouse has come home from work. Spontaneous beliefs arise from processes of mnemonic elaboration. A representation of the stimulus, e.g., the noise coming from downstairs, is used as a cue to retrieve related material about causes and consequences from diverse types of long-term mnemonic/conceptual knowledge stores (Uleman et al., 2008). (As we shall see in §5.2.1, there is good evidence that people can intervene by means of memory control actions to arrest or redirect this process of mnemonic elaboration.) A similar picture applies to evaluations; people routinely assess the goodness or badness of objects and situations they confront in an ongoing way, and they usually do so spontaneously, without extensive deliberation. Emotions produce characteristic biases on these spontaneous belief and evaluation formation mechanisms. For example, people experiencing occurrent anxiety interpret ambiguous stimuli as threatening and evaluate them more negatively (Eysenck, Mogg, May, Richards, & Mathews, 1991), and corresponding results are observed for other emotions (Bower, 1991).

Memory/Thought Mechanisms – Emotions bias patterns of memory retrieval such that emotion-congruent memory items are rendered more accessible and spontaneously emerge, or even intrude, into consciousness (Bower & Cohen, 2014; LeDoux, 1993). Additionally, one's spontaneous thoughts—including mind wandering, daydreams, and related types of thought—are biased towards emotion-congruent material (Smallwood, Fitzgerald, Miles, & Phillips, 2009).

4.3. Emotions Serve as "Base States" for Response Pulses

I now want to clarify the connection between emotion-produced biases and the kinds of spontaneous states introduced in my discussion of cognitive control, response pulses. To do this, it will be useful to bring in a new bit of terminology. Earlier in discussing conflict tasks, I described specific response pulses that are activated in these tasks during specific stimulus conditions. For example, in the Stroop task, a response pulse to read words is produced when word stimuli are presented. One might ask: Why does the relevant mechanism produce *this* response pulse in this stimulus condition, rather than some other response pulse directed at some other response? The answer appeals to the presence of an acquired habit: because word reading is extensively practiced over the course of years and decades, a word reading habit is formed.

A habit is an example of what we can call a *base state*, a state that provides an answer to the preceding type of question regarding why one response pulse is elicited rather than another. More precisely, given that a mechanism *M* produces a response pulse to *R* in situation *S*, a base state explains why *M* produces this particular response pulse to *R* in *S* rather than some other response pulse.

The notion of a base state helps in characterizing the effects of emotions on our psychology: In virtue of the biases that emotions produce on diverse psychological mechanisms, emotions too serve as base states for response pulse. That is, because of these biases, diverse psychological mechanisms now produce emotion-congruent response pulses during the interval in which the emotion is active. Consider a bias on attention during fear. In virtue of this bias, even very weakly threat-relevant stimuli now evoke response pulses to shift attention. Or consider a bias on memory during sadness. In virtue of the bias, otherwise ordinary stimuli in the environment evoke response pulses to retrieve negative associated memories.

4.4. Other Emotion-Type States Also Serve as Base States for Response Pulses

The preceding claim about emotions—that they are base states for response pulses —applies to other emotion-type states as well. This follows from the fact that other emotion-type states also produce biases across diverse psychological mechanisms. A key difference is that while emotions generate the widest profile of biases across psychological systems, other emotion-type states typically affect a somewhat narrower range of psychological systems.

Consider pains, itches, and feelings of fatigue. These states produce robust biases on action selection. In particular, they produce action tendencies to, respectively, withdraw from the source of pain, scratch the itchy area, and stop the fatiguing activity. They also have profound effects on attention, sometimes making it difficult for the person to entertain other thought contents. But it is not clear they have some of the other characteristic effects of emotions, for example on belief formation, memory, or spontaneous thought, where these effects go beyond the powerful effects these states have on attention. Cravings (for example, for drugs), drives (such as hunger and thirst), and impulses (such as hair pulling impulses seen in the psychiatric disorder trichotillomania) are likely somewhere in between emotions on the one hand and pains/itches/fatigue on the other hand in the breadth of psychological mechanisms that they bias.[18]

4.5. The Response Pulses Produced by Emotion-Type States Influence Action

Thus far I have argued that emotion-type states are base states for temporally-extended streams of response pulses. I now want to focus on the consequences of these response pulses for action. Specifically, I want to show that response pulses strongly favor the production of actions congruent with the emotion-type state.

Response pulses influence action through a pathway that involves decisions: response pulses influence one's decisions and one's decisions, in turn, bring about overt action. To see this overall pathway at work, it will be useful to have a case in mind. Consider a man who is humiliated by his boss and co-workers at a company meeting, and he retreats to his cubicle seething with anger. During the extended interval that the emotion is active, he experiences ongoing streams of activated response pulses that influence what he notices as well as his evaluations, goals, recollections, and thoughts. Here are some of them. He has spontaneous shifts in attention: He keeps noticing his co-workers' voices while he is trying to work. Their voices are now highly salient and he spontaneously attends to them each time they speak. His spontaneous evaluations and goals change: The prospect of telling his boss off, which he would have previously have found unthinkable, now strikes him as deeply satisfying. His spontaneous memories change: He keeps recalling the meeting where he was humiliated; it plays in his mind over and over again. His spontaneous thoughts change: He is trying to read and understand an important email, but his thoughts keep turning to fantasies about getting back at his boss.

The preceding consequences of the man's response pulses—consequences on what he notices, how he evaluates prospects, what he recalls, and what thoughts he thinks—will plausibly affect

[18] More detailed discussions of some of these emotion-type states can be found in the following works: pain (Klein, 2015), cravings (Skinner & Aubin, 2010), urges in trichotillomania (Madjar & Sripada, 2016).

the man's decisions through multiple routes. Some routes are more immediate: a spontaneous tendency to evaluate an action positively (e.g., seeing telling off your boss as deeply satisfying) will tend to directly bias decisions in favor of that action. Other routes are mediated by further mental activity. For example, when a person's attention, recollections, and thoughts are constantly redirected to a certain negative event, this will in turn affect new episodes of belief formation and evaluation (typically, new beliefs and evaluations are formed that greatly exaggerate the severity and aversiveness of the event). These new beliefs and evaluations will in turn strongly bias the person's decisions.

Overall, then, in the picture I have put forward, response pulses are the motivational "tips of the spear" of emotion-type states. Emotion-type states produce temporally-extended streams of response pulses across diverse psychological mechanisms, and these response pulses in turn strongly bias action.[19]

5. The Atomic Model of Self-Control

5.1. Bridging Cognitive Control and Self-Control

Three key pieces for my account of self-control are now in place. First, we have an account of response pulses: simple, brief states that dispose the person to responses involving attention, belief, memory, thought, and action. Second, we have an account of cognitive control actions that target response pulses. And third, we have an account of emotion-type states that makes clear that they are sources of temporally-extended streams of response pulses across multiple psychological mechanisms and that these response pulses in turn influence action.

If all of this is right, then a bridge between cognitive control and self-control starts to come into focus. The main idea behind this bridge is that there is an intriguing componential relationship between self-control and cognitive control, and I now want to lay it out explicitly.

> **Atomic Model of Self-Control**: Exercises of self-control consist of skilled sequences of cognitive control aimed at regulating the temporally-extended streams of response pulses associated with an emotion-type state, in order to prevent the emotion-type state from being effective in action.

[19] A further claim is that response pulses are the *only* pathway by which emotion-type states influence action. I believe this further claim is in fact correct, but I leave its defense for another day.

In short, the atomic model says that self-control and cognitive control aren't qualitatively different—they aren't produced by distinct mental systems or brain mechanisms—but rather they are "quantitatively" different. Engaging in self-control consists in engaging in numerous exercises of cognitive control over time to deal with the temporally-extended streams of response pulses produced by emotion-type states, and which are a major way these states influence action.[20]

Notice the atomic model says that the constituent exercises of cognitive control that make up an exercise of self-control must be *skilled*. The reason to insist on this qualifier is that keeping an emotion-type state from being effective in action requires suppressing temporally-extended streams of response pulses across multiple psychological domains, i.e., action selection, attention, belief, evaluation, memory, and thought. This is a complex endeavor that requires performing the right cognitive control actions at the right time with the right intensity for the right duration. Thus an exercise of self-control doesn't consist of just any arbitrary sequence of cognitive control actions. It is rather a sequence that manifests the appropriate sort of *knowing-how* to block the actional upshots of an emotion-type state.

5.2. Evidence for the Atomic Model of Self-Control

There are a number of lines of evidence that support the atomic model, and I want to now summarize a few.

5.2.1. Argument from Actual Self-Control Strategies

[20] There is a need to clarify one feature of the atomic model that arises due to ambiguity in talking about cognitive control. If I exercise cognitive control to get the correct answer on a trial of the Stroop task, getting the correct answer is not the exercise of cognitive control. Rather it is *what is enabled* because I exercised cognitive control against the response pulse to read words. Were it not for this "direct" exercise of control against the response pulse, getting the correct answer would not have happened, as the response pulse would have manifested in action instead. Getting the right answer is thus what is called in the literature a "controlled action". Given this distinction, some elements of self-control might consist of direct exercises of cognitive control, say redirecting attention *away* from a temptation that is currently grabbing one's attention. Other elements of self-control are the controlled actions so enabled by such direct exercises of control, such as redirecting attention *towards* one's reasons to stay on one's diet. Because direct exercises of cognitive control and the controlled actions so enabled are tightly fused in most cases, I usually leave out the qualification going forward.

One line of evidence for the atomic model comes from looking at studies that have attempted to examine mechanistically how people actually regulate emotion-type states. Most of these studies focus on regulation of emotions (Gross, 1998), though the strategies appear to be applicable more broadly. These studies identify a small set of strategies that are commonly used (Naragon-Gainey, McMahon, & Chacko, 2017), and when we examine these strategies closely, they appear to critically involve cognitive control.

Emotion regulation strategies are typically classified as response-focused or antecedent-focused (Gross, 1998). One important response-focused emotion regulation strategy is called "expressive suppression", which includes suppressing the facial gestures, postural changes, approach/withdrawal tendencies (for example, the tendency to flee during fear), and other elements of an emotion's profile of action tendencies (Gross & Levenson, 1993). This strategy is naturally understood as involving the inhibitional family of cognitive control actions.

The other two emotion regulation strategies I'll discuss are antecedent-focused; they regulate states involved in emotion elicitation. One strategy is distraction. Stimuli that evoke and/or maintain emotions are salient and grab attention, and distraction involves volitionally redirecting attention away from these salient stimuli. As such, this strategy is naturally understood in terms of cognitive control, specifically the attentional family of cognitive control actions if the salient stimulus is in the external environment, or the memory/thought suppression family of control actions if the salient stimulus is an occurrent memory or thought.

Another antecedent-focused emotion regulation strategy, and the most widely studied, is reappraisal. Earlier, I noted that emotions arise and are maintained due to appraisals, spontaneous interpretations in which the current situation is assigned a meaning with respect to one's standing concerns (Ellsworth & Scherer, 2003; Scherer, 2001). Importantly, there is substantial evidence that appraisal processes rely on ongoing spontaneous retrieval from semantic and episodic memory; this mnemonic information is needed to contextualize a stimulus and assess its self-relevance (see Ochsner and Feldman Barrett, 2001 for a review). The goal of reappraisal is to override ongoing spontaneous appraisals and replace them with alternatives that are less conducive to maintenance of the emotion.

Notice that on this picture of how appraisal processes work, appraisal states are themselves dependent on response pulses associated with a number of psychological mechanisms, such as memory retrieval mechanisms, and thus regulating appraisal states via reappraisal is a form of cognitive control. This claim in turn implies that we should be able to identify specific control actions associated with reappraisal. This is indeed the case. Evidence is emerging that memory suppression-type control actions are a critical element of reappraisal—the person suppresses

mnemonic material associated with spontaneous interpretation of events enabling them to voluntarily generate alternatives (Engen & Anderson, 2018). So here again, analysis of the mechanisms of emotion regulation finds clear evidence that its constituent elements are exercises of cognitive control.

When we turn from emotions to other kinds of emotion-type states, analogues of the preceding three strategies are apparent. Consider for example cravings for unhealthy foods. Strategies to deal with such cravings include expressive suppression (for example, inhibiting the urge to reach once again into the bag of chips), distraction (e.g., looking away from the tempting item and thinking about other things), and reappraisal (e.g. thinking of marshmallows as fluffy clouds or thinking of cookies as ugly lumps of fat).[21] Taken together, these observations suggest that the atomic model is right that exercises of self-control are built out of constituent exercises of cognitive control.

5.2.2. Argument from Things We Can (And Cannot) Control

A second argument for the atomic model is based on close observation of what aspects of emotion-type states we can and cannot control. Here is a very general and highly underappreciated feature of how cognitive control works:

> **(Limit - Cognitive Control)** Cognitive control is limited to regulating response pulses, and it cannot (directly) regulate the base states that produce them.

We can vividly see this restriction at work in the case of the Stroop task. A person can certainly exercise cognitive control against activated response pulses for the word reading response, which is why they get the right answer on incongruent trials most of the time. But they cannot, no matter how hard they try, ever exercise cognitive control over the base state (i.e., the underlying word reading habit) in virtue of which these activated response pulses arise. This is why even after hundreds of trials of the Stroop task, the word reading response pulse arises trial after trial, and must be suppressed on every single occasion.

With the preceding limit in mind, I now want to consider an interesting and underappreciated question: When we regulate emotion-type states with an eye to preventing these states from

[21] Strategies such as these are discussed by Mischel and colleagues in classic work on delay of gratification in children, see Mischel and Moore, 1980; Mischel and Mischel, 1987.

influencing action, which of the components shown in Figure 1 can *directly* be the targets of control and which cannot? I believe that if we reflect on this question, it becomes clear that the only things over which we have direct control are the response pulses associated with the emotion-type state (shown in green). With regard to all the other elements in the figure, we do not have direct control over them. Let me elaborate on this claim.

I have already argued that people can regulate the temporally-extended streams of response pulses produced as a consequence of the activation of emotion-type states (green box on the right in Figure 1). I talked about the cognitive control research program, which uses conflict tasks to elucidate the mechanisms by which we control response pulses associated with action, attention, memory, and thought. I also considered concrete cases such as the man in his cubicle. Each time his attention inappropriately shifts, he can bring it back; each time his thoughts inappropriately stray, he can redirect them; each time he has urges to tell off his boss, he can inhibit them.

I also already argued that people can often regulate the appraisal process (green box on the left in Figure 1) that elicits and sustains ongoing emotion-type state episodes. In particular, in §5.2.1, I argued that appraisals are dependent on response pulses produced by various psychological mechanisms (such as memory retrieval mechanisms), and we can often regulate appraisals, at least to a limited extent, through memory/thought control-type actions.

When we turn to the other boxes box in Figure 1, i.e. the box for the activated emotion-type state and their associated biases on psychological mechanisms, it appears we cannot directly regulate these. Consider again the man in his cubicle. He has biases across multiple mechanisms that produce altered temporally-extended streams of response pulses. But while the man can directly control each individual response pulse, he cannot *directly* control the biases themselves. So for example, while he can directly regulate each individual response pulse to shift attention that he experiences, he cannot directly regulate the underlying biases of his attention mechanisms that produce the attention-redirecting response pulses in the first place.

Now, to be clear, we are not powerless over the biases produced by emotion-type states; we do have a certain measure of *indirect* control over them. Such indirect control is achieved by altering the appraisal processes by which the relevant emotion-type state is elicited and sustained, which has the effect of attenuating all the downstream consequences of the emotion-type state.

The preceding observations suggest the following general claim:

(**Limit - Self-Control**) In controlling an emotion-type state to prevent it from being effective in action, the only things we can *directly* control are the response pulses that are involved in the state's elicitation or the response pulses produced as a consequence of the state's elicitation.

If **Limit – Self-Control** is correct, then this makes a strong case that self-control consists of cognitive control based on two complementary arguments. The first argument starts by noting that cognitive control is the means by which we exercise control over activated response pulses arising from diverse psychological mechanisms. If the only things we can directly control about an emotion-type state are its associated response pulses, then control over emotion-type states must then be cognitive control.

The second argument is based on asking the deeper question of *why* **Limit – Self-Control** is true. The best explanation seems to be that self-control consists of cognitive control. That is, it is hoped that most readers are already convinced that cognitive control is limited to regulating activated response pulses, but not the base states that produce them. If self-control consists of cognitive control, this would explain why it too must obey a broadly analogous restriction.

5.2.3. Argument from Correlated Abilities

An important prediction of the atomic model is that people who are worse at cognitive control, as measured by conflict tasks, will do worse on measures of self-control. There is a sizable body of evidence that this is in fact the case that is derived from studies of patients with psychiatric disorders (Lipszyc & Schachar, 2010; Schachar, Tannock, & Logan, 1993). For example, large meta-analyses find individuals with attention-deficit/hyperactivity disorder (ADHD), a disorder characterized by deficits in self-control, score substantially lower than typical individuals in a number of conflict tasks (Frazier, Demaree, & Youngstrom, 2004; Huang-Pollock, Karalunas, Tam, & Moore, 2012; Willcutt, Doyle, Nigg, Faraone, & Pennington, 2005). When we look at populations without psychiatric disorders, correlations between performance on conflict tasks and measures of self-control are still observed, though they tend to be more modest in size (Duckworth & Kern, 2011).

One likely explanation for the differences across the populations that are studied is that standard conflict tasks are relatively easy—they have to be to get consistent results across dozens or hundreds of nearly identical trials, which is what is needed for current methods of statistical analysis of these tasks. As a result, these tasks either don't measure variation in the typical range effectively. Or, alternatively, the variation they do measure in the typical range is

mostly irrelevant to having poor self-control outcomes in the real world. That is, measurable problems with self-control in day-to-day life only arise when levels of cognitive control, as measured by standard conflict tasks, are *substantially* below the mean.

Putting these complexities aside, the main point I want to emphasize is that the atomic model makes what is on its face an unlikely prediction: A person's performance on a set of highly structured, repetitive cognitive tasks in the laboratory, such as the Stroop task and Go/No Go task, will be predictive of how they perform at self-control in the real world when confronted with temporally extended, affectively charged, complex states such as emotions and cravings. This prediction of the atomic model is in fact supported by the weight of evidence.

Part 6. Self-Control: Theoretical and Commonsense Conceptions

The atomic model provides an attractive theoretical account of self-control. A chief advantage of the view, which I presently discuss, is that it shows why self-control constitutes a principled, well-behaved theoretical kind. But of course self-control is not a notion introduced *de novo* by scientific theorizing. There is an antecedent notion of self-control already present in common sense, which is reflected in ordinary judgments of about what does and does not count as self-control. What is the relationship between the atomic model's account of self-control and the commonsense conception? In this final part of the paper, I argue that the two are in tight alignment. In making my case, my strategy is to contrast the atomic account with a family of popular views of self-control called "results" views. I then show there is a key ingredient in the commonsense conception of self-control that is missing from results views, and the atomic model is well-positioned to capture it.

6.1. Self-Control as a Unified Theoretical Kind

At first pass, self-control seems to encompass a motley assortment of things. People exercise self-control in many different ways, for example: by directly stopping themselves from acting on an urge, by redirecting their attention away from a tempting object, by effortfully conjuring up vivid images of the negative consequences of an action, and through various combinations of the preceding, among other ways. They do so over very different timescales: Resisting a momentary itch versus resisting the urge to mind wander during an all-day seminar. Additionally, they direct self-control at disparate target states: emotions of various kinds, food or drug cravings,

drives such as hunger and thirst, impulses, pain, fatigue, and so on. What makes these exercises all of the same kind?[22]

The atomic model says there is in fact underlying unity here, but to appreciate it, we need to shift to a lower-level perspective where we consider elemental units. Consider the case of chemistry, where an understanding of the relevant elemental units helps to make sense of why certain things that superficially appear to differ (e.g., coal, graphite, diamonds) are in fact members of a unified kind (allotropes of carbon). Similarly, the atomic model says self-control exercised in all the preceding ways against all the preceding targets is unified because at a more basic level it always consists of just one thing: sequences of cognitive control actions that proximally target activated response pulses, and that ultimately have the aim of blocking the actional upshots of emotion-type states.

6.2. Process Versus Results Views of Self-Control

There is an ongoing controversy among theorists in philosophy and cognitive science about how to understand self-control. Broadly speaking, two families of views can be discerned (Duckworth, Gendler, & Gross, 2016). First there are "process" views that say when you exercise self-control, you engage a distinctive type of mental process. It is common to cast this special intra-psychic doing in vague terms, often with the aid of metaphors ("resisting", "reining in", "subduing" a desire). The atomic model gives substantial new resources to the process approach. It lets us fill in these metaphors with detailed micro-level theories involving cognitive control actions, response pulses, constituent structure of emotion-type states, and so on.

The leading alternatives are "results" views of self-control. These views put the emphasis on the outcome that self-control, if it is successful, brings about, leaving it essentially open what process is used to reach the outcome. In psychology, one commonly encountered results view says self-control consists in choosing a larger, later reward over a smaller, earlier one. In philosophy, an influential version of a results view is put forward by Al Mele (Mele, 1987, 2003). On his view, the relevant result that defines self-control is mastery over desires that are contrary to one's all things considered best judgments. On both of these views, so long as what a

[22] In an important recent article, Marcela Herdova (2017) brings careful scrutiny to the question of whether self-control constitutes a well-behaved, cohesive theoretical kind. Her criticisms are primarily directed at Al Mele's "neo-Aristotelian" account of self-control, and I consider related criticisms in §6.3. The atomic model I put forward can be seen as a response to Herdova's broader challenge to theorists to provide an account of self-control that shows why it is a cohesive theoretical kind.

person does appropriately brings about the relevant result (or attempts to do so), then what they do counts as self-control.[23]

When we look more closely, however, results views encounter serious counter-examples. These counterexamples highlight that a key ingredient for self-control seems to be missing from results views. A bit later, I argue that the atomic account is well-positioned to capture this missing ingredient.

6.3. Problems for Results Views

I want to briefly present a few counterexamples to results views. My focus here is on Mele's "mastery of motivation" view, but my remarks can be readily adapted to other versions of the results approach.

> (**A**) Bo has a powerful itch on his forearm. He judges it is best that he not scratch it or else it will scar, so he puts some calamine lotion on it and the urge to scratch immediately goes away.

It is clear that Bo masters motivation that is contrary to his best judgment. However, what he does is clearly not an exercise of self-control.

> (**B**) Jo is claustrophobic and can't get herself to enter a crowded theater, even though she judges it is best to enter (suppose it would mean a lot to her son). Luckily, she is also terribly phobic of clowns so she hires Bozo to chase her in.[24]

Jo too clearly masters motivation that is contrary to her best judgment. Once again, however, what she does isn't an exercise of self-control.

---

[23] Some views of self-control emphasize strategically selecting one's environment in a way that prevents the elicitation of problematic impulses (for example, Duckworth, Gendler and Gross, 2016). These views also fall under the results family of views.

[24] I thank Jesse Summers for inspiring this example.

Consider another case based on the distinction made by many theorists between synchronic and diachronic forms of self-control. Self-control is synchronic when it is directed at a desire that is currently active, and it is diachronic when it prevents an unwanted, non-occurrent, anticipated future desire from becoming active.[25] Though this distinction is widely used, I believe that when we look closely at cases of diachronic self-control, they look a lot like cases **A** and **B**. Here is an example:

> (**C**) Mo judges it is best that he not smoke, so he visits a doctor who specializes in helping smokers quit quickly and effortlessly. The doctor hands Mo a special little pill and tells him it tastes like candy. Mo has no desire to smoke at the moment, but he knows he will have one shortly. He takes the pill, and he never has a desire to smoke again.

What Mo does in this case certainly fits the definition of "diachronic self-control"—he prevents an unwanted, non-occurrent, anticipated future desires from becoming active. But just like cases **A** and **B**, what Mo does clearly does not involve exercising self-control.

This conclusion is reinforced when we contrast **C** with a synchronic case:

> (**D**) Ro judges it is best that she not smoke. However, right now, she has a very strong desire to smoke. She immediately effortfully resists acting on this desire until it passes. As a result, she does not smoke.

Here there is no doubt that what Ro does is an instance of self-control.[26] Overall, then, results views seem to get things wrong, and in the next section I make a proposal about what specifically is missing from these views.

---

[25] See Kennett and Smith, 1996 and Smith, 2001.

[26] If cases like **C** pretty clearly don't count as self-control while cases like **D** clearly do, why have many theorists been attracted to the notion of diachronic self-control in the first place? The explanation, I believe, is that **C** is notable in that it involves "pure" diachronic self-control. Theorists tend to ignore such cases and instead focus on impure cases in which both diachronic strategies for preventing future desires are co-present with effortful synchronic regulation. Here is an example:

> (**E**) Wo judges it is best that she not smoke. Her plan is to throw her full carton of cigarettes into the incinerator. Wo doesn't have an urge to smoke now, but when she tries to let go of the carton,

6.4. Mental Effort and the Vindication of Common Sense

I propose that the ordinary notion of self-control restricts what counts as self-control to the performance of a certain set of intra-psychic actions that have a distinctive phenomenology involving mental effort[27]. The folk cannot say with any precision what is the nature of these intra-psychic actions. But the feeling of effort that accompanies performing them is quite salient and that gets incorporated as a central element of their self-control concept. So the folk understanding of exercises of self-control might be characterized as "the effortful mental stuff I do to prevent myself from acting on emotions, cravings, impulses, and similar states." If this proposal is right, we get a clear explanation for two things: why common sense sharply diverges from results views of self-control, and why it remains tightly coupled to the atomic view.

Divergence between common sense and the results view arises because this view sets no constraints on the means by which the relevant results are brought about. Whenever these means bypass effortful regulation, as in **A**, **B**, and **C**, the results view will count as an instance of self-control something that the folk plainly refuse to classify as such.

The alignment between ordinary judgment and the atomic model arises because this model says self-control consists of sequences of cognitive control—these are the atoms of self-control. But critically, cognitive control is tightly linked with mental effort. This can be appreciated just by looking at the conflict tasks: all the incongruent conditions that involve exercising cognitive control to suppress response pulses feel effortful, while all of the congruent conditions where subjects act on response pulses (and do not exercise cognitive control) are felt as relatively effortless.

> she has a sense of dread and hesitates—she can't bring herself to do it. She knows in an hour she will have a strong urge to smoke, and, since she has no money, she has no way to buy a new pack. She anticipates how miserable she will feel if her future urges to smoke go unsatisfied. Nonetheless, Wo effortfully inhibits her anxiety and dread and drops the carton into the incinerator.

What Wo does in this case *does* count as an exercise of self-control. But the difference-maker in this case in comparison to **C** appears to be the presence of effortful synchronic regulation.

[27] For a discussion of effort and self-control, see Holton, 2009, esp. chap. 6. For general discussions of the role of the phenomenology of effort in agency, see Bayne and Levy, 2006 and Brent, 2017.

But effort and cognitive control don't just happen to be correlated without any further explanation. Rather, the experience of effort during cognitive control appears to be a deep feature of how it works. According to an influential recent model (Shenhav et al., 2017), the experience of effort arises from the "cost of control representation" that is an essential component of EVC calculation, which in turn is the basis by which cognitive control actions are selected and sequenced. Putting a philosophical gloss on this model, effort serves as a non-conceptual valence-type representation (Carruthers, 2017), with a distinctive phenomenal character, of the disvalue of exercising cognitive control in a particular context.[28]

If the folk notion of self-control centrally involves the idea of mental effort and if the exercise of cognitive control is inherently effortful, then we have an explanation of why the atomic model of self-control is closely aligned with the folk notion. The folk don't know much about the atoms of self-control, i.e., cognitive control actions that target response pulses. These are things revealed to us by cognitive science, for example through the careful study of conflict tasks. But the folk do have experiences of the effortful phenomenology of cognitive control actions, and thus their naïve verdicts about self-control and the verdicts of the atomic model remain in tight harmony. This is an intriguing case where empirical science vindicates common sense, with phenomenology providing the basis of the vindication.

7. Conclusion

What happens when a person exercises self-control in order to regulate states such emotions or cravings? In this paper, I proposed the atomic account, which draws heavily on the sizable research program in cognitive science on cognitive control. Self-control targets states such as emotions and cravings, which are complex and temporally-extended (typically lasting minutes to hours). Cognitive control, however, targets something else: simple states called response pulses that unfold over hundreds of milliseconds. Though the two kinds of control have very different targets, I argued that they are nonetheless intimately connected: they are related as part to whole. According to my account, exercises of self-control consist in performing numerous cognitive control actions in a skilled way over time. These cognitive control actions, in turn, target temporally-extended streams of response pulses that are produced by emotions, cravings, and similar states, thus preventing these states from being effective in action.

On the atomic account, cognitive control actions are the atoms of self-control, the elemental units that get combined in complex ways to produce different kinds of exercises of self-control.

[28] The standard view, which I think is broadly on the right track, is that disvalue attaching to exercises of cognitive control typically arises from opportunity costs: the machinery that subserves cognitive control could be deployed for other useful purposes (see Kurzban *et al.*, 2013).

Somewhat surprisingly, the boundaries of self-control that emerge from this account turn out to align quite nicely with the boundaries drawn by common sense, and it was argued that the phenomenology of mental effort helps to explain this correspondence.

Acknowledgements

Work on this manuscript was supported by a grant from the John Templeton Foundation Philosophy and Science of Self-Control Project. Thanks to Hannah Altehenger, Juan Pablo Burmúdez, Sarah Buss, and Alfred Mele for comments on earlier versions of the manuscript. An anonymous reviewer at *Noûs* provided extensive comments and suggestions that greatly improved the manuscript. Thanks to audiences at the Summer Seminars in Neuroscience and Philosophy at Duke University 2017, Philosophy and Science of Self-Control Summer School at University of Florida, Tallahassee, Florida 2017, the Getting Better at Simple Things Workshop in Bogotá, Columbia 2018, and the Philosophical Moral Psychology Workshop, Munich, Germany 2018.

References

Anderson, M. C., & Green, C. (2001). Suppressing unwanted memories by executive control. *Nature*, *410*, 366.

Anderson, M. C., Ochsner, K. N., Kuhl, B., Cooper, J., Robertson, E., Gabrieli, S. W., … Gabrieli, J. D. (2004). Neural systems underlying the suppression of unwanted memories. *Science*, *303*, 232–235.

Andrews-Hanna, J. R., Smallwood, J., & Spreng, R. N. (2014). The default network and self-generated thought: Component processes, dynamic control, and clinical relevance. *Annals of the New York Academy of Sciences*, *1316*, 29–52.

Aron, A. R., Robbins, T. W., & Poldrack, R. A. (2004). Inhibition and the right inferior frontal cortex. *Trends in Cognitive Sciences*, *8*, 170–177.

Aron, A. R., Robbins, T. W., & Poldrack, R. A. (2014). Inhibition and the right inferior frontal cortex: One decade on. *Trends in Cognitive Sciences*, *18*, 177–185.

Baddeley, A. D. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology*, *63*, 1–29.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In *Psychology of learning and motivation* (Vol. 8, pp. 47–89). Elsevier.

Bayne, T. J., & Levy, N. (2006). The Feeling of Doing: Deconstructing the Phenomenology of Agency. In *Disorders of Volition* (pp. 49–68). Cambridge, MA: MIT Press.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychol Rev*, *108*, 624–652.

Botvinick, M. M., & Cohen, J. D. (2014). The Computational and Neural Basis of Cognitive Control: Charted Territory and New Frontiers. *Cognitive Science*, *38*, 1249–1285.

Botvinick, M. M., Cohen, J. D., & Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: An update. *Trends in Cognitive Sciences*, *8*, 539–546.

Bower, G. H. (1991). Mood congruity of social judgments. *Emotion and Social Judgments*, 31–53.

Bower, G. H., & Cohen, P. R. (2014). Emotional influences in memory and thinking: Data and theory. *Affect and Cognition*, *13*, 291–331.

Brent, M. (2017). Agent causation as a solution to the problem of action. *Canadian Journal of Philosophy*, *47*, 656–673.

Carruthers, P. (2015). *The centered mind: What the science of working memory shows us about the nature of human thought*. OUP Oxford.

Carruthers, P. (2017). Valence and value. *Philosophy and Phenomenological Research*.

Christoff, K., Irving, Z. C., Fox, K. C., Spreng, R. N., & Andrews-Hanna, J. R. (2016). Mind-wandering as spontaneous thought: A dynamic framework. *Nature Reviews Neuroscience*, *17*, 718.

Cohen, J. D. (2017). Cognitive Control. In *The Wiley Handbook of Cognitive Control* (pp. 1–28). Wiley-Blackwell.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nat Rev Neurosci*, *3*, 201–215.

Davidson, D. (1963). Actions, Reasons, and Causes. *The Journal of Philosophy*, *60*, 685–700.

Deussen, P. (1980). *Sixty Upaniṣads of the Veda* (Vol. 1). Motilal Banarsidass Publ.

Dill, B., & Holton, R. (2014). The Addict in Us all. *Frontiers in Psychiatry*, *5*, 1–20.

Donders, F. C. (1969). On the speed of mental processes. *Acta Psychologica*, *30*, 412–431.

Duckworth, A. L., Gendler, T. S., & Gross, J. J. (2016). Situational strategies for self-control. *Perspectives on Psychological Science*, *11*, 35–55.

Duckworth, A. L., & Kern, M. L. (2011). A meta-analysis of the convergent validity of self-control measures. *Journal of Research in Personality*, *45*, 259–268.

Egner, T., & Hirsch, J. (2005). Cognitive control mechanisms resolve conflict through cortical amplification of task-relevant information. *Nature Neuroscience*, *8*, nn1594.

EinhÃ, W., Rutishauser, U., & Koch, C. (2008). Task-demands can immediately reverse the effects of sensory-driven saliency in complex visual stimuli. *Journal of Vision*, *8*, 2–2.

Ellsworth, P. C., & Scherer, K. R. (2003). Appraisal processes in emotion. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Handbook of affective sciences* (pp. 572–595). New York, NY, US: Oxford University Press.

Engen, H. G., & Anderson, M. C. (2018). Memory Control: A Fundamental Mechanism of Emotion Regulation. *Trends in Cognitive Sciences*.

Evans, J. S. B. (2007). On the resolution of conflict in dual process theories of reasoning. *Thinking & Reasoning*, *13*, 321–339.

Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-Process Theories of Higher Cognition Advancing the Debate. *Perspectives on Psychological Science*, *8*, 223–241.

Eysenck, M. W., Mogg, K., May, J., Richards, A., & Mathews, A. (1991). Bias in interpretation of ambiguous sentences related to threat in anxiety. *Journal of Abnormal Psychology*, *100*, 144–150.

Forstmann, B. U., Ratcliff, R., & Wagenmakers, E.-J. (2016). Sequential sampling models in cognitive neuroscience: Advantages, applications, and extensions. *Annual Review of Psychology*, *67*.

Frazier, T. W., Demaree, H. A., & Youngstrom, E. A. (2004). Meta-analysis of intellectual and neuropsychological test performance in attention-deficit/hyperactivity disorder. *Neuropsychology*, *18*, 543–555.

Frijda, N. H. (1986). *The emotions*. Cambridge University Press.

Frijda, N. H. (1987). Emotion, cognitive structure, and action tendency. *Cognition and Emotion*, *1*, 115–143.

Frijda, N. H., Kuipers, P., & Ter Schure, E. (1989). Relations among emotion, appraisal, and emotional action readiness. *Journal of Personality and Social Psychology*, *57*, 212.

Gagnepain, P., Hulbert, J., & Anderson, M. C. (2017). Parallel regulation of memory and emotion supports the suppression of intrusive memories. *Journal of Neuroscience*, 2732–16.

Gross, J. J. (1998). The Emerging Field of Emotion Regulation: An Integrative Review. *Review of General Psychology*, *2*, 271–299.

Gross, J. J., & Levenson, R. W. (1993). Emotional suppression: Physiology, self-report, and expressive behavior. *J Pers Soc Psychol*, *64*, 970–986.

Henden, E. (2008). What is self-control? *Philosophical Psychology*, *21*, 69–90.

Henden, E., Melberg, H.-O., & Rogeberg, O. (2013). Addiction: Choice or Compulsion? *Frontiers in Psychiatry*, *4*. https://doi.org/10.3389/fpsyt.2013.00077

Herdova, M. (2017). Self-control and mechanisms of behavior: Why self-control is not a natural mental kind. *Philosophical Psychology*, *30*, 731–762.

Hofmann, W., Schmeichel, B. J., & Baddeley, A. D. (2012). Executive functions and self-regulation. *Trends in Cognitive Sciences*, *16*, 174–180.

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

Holton, R., & Berridge, K. (2013). *Addiction between compulsion and choice* (N. Levy, Ed.). OUP USA.

Huang-Pollock, C. L., Karalunas, S. L., Tam, H., & Moore, A. N. (2012). Evaluating vigilance deficits in ADHD: A meta-analysis of CPT performance. *Journal of Abnormal Psychology*, *121*, 360–371.

Itti, L., & Borji, A. (2015). Computational models: Bottom-up and top-down aspects. *ArXiv Preprint ArXiv:1510.07748*.

Kennett, J. (2003). *Agency and responsibility: A common-sense moral psychology*. Oxford: Clarendon.

Kennett, J., & Smith, M. (1996). Frog and Toad Lose Control. *Analysis*, *56*, 63–73.

Klein, C. (2015). *What the body commands: The imperative theory of pain*. MIT Press.

Kool, W., Shenhav, A., & Botvinick, M. M. (2017). Cognitive Control as Cost-Benefit Decision Making. In *The Wiley Handbook of Cognitive Control* (pp. 167–189). Wiley-Blackwell.

Kurzban, R., Duckworth, A., Kable, J. W., & Myers, J. (2013). An opportunity cost model of subjective effort and task performance. *The Behavioral and Brain Sciences*, *36*, 661–679.

LeDoux, J. E. (1993). Emotional memory systems in the brain. *Behavioural Brain Research*, *58*, 69–79.

Levenson, R. W. (1994). Human emotion: A functional view. *The Nature of Emotion: Fundamental Questions*, *1*, 123–126.

Levy, N. (2011). Resisting 'Weakness of the Will.' *Philosophy and Phenomenological Research*, *LXXXII*, 134–155.

Levy, N. (Ed.). (2013). *Addiction and Self-Control: Perspectives from Philosophy, Psychology, and Neuroscience* (1 edition). New York, NY: Oxford University Press.

Lipszyc, J., & Schachar, R. (2010). Inhibitory control and psychopathology: A meta-analysis of studies using the stop signal task. *Journal of the International Neuropsychological Society*, *16*, 1064–1076.

Loewenstein, G. (1996). Out of control: Visceral influences on behavior. *Organizational Behavior and Human Decision Processes*, *65*, 272–292.

Madjar, S., & Sripada, C. S. (2016). The Phenomenology of Hair Pulling Urges in Trichotillomania: A Comparative Approach. *Frontiers in Psychology*, *7*, 199.

McKoon, G., & Ratcliff, R. (1992). Inference during reading. *Psychological Review*, *99*, 440.

Mele, A. R. (1987). *Irrationality: An Essay on Akrasia, Self-Deception, and Self-Control*. New York: Oxford Univiverity Press.

Mele, A. R. (1997). Underestimating Self-control: Kennett and Smith on Frog and Toad. *Analysis*, *57*, 119–123.

Mele, A. R. (2003). *Motivation and Agency*. Oxford University Press, USA.

Mele, A. R. (2013). Self-control, motivational strength, and exposure therapy. *Philosophical Studies*, *170*, 359–375.

Melloni, L., Leeuwen, S. van, Alink, A., & Müller, N. G. (2012). Interaction between Bottom-up Saliency and Top-down Control: How Saliency Maps Are Created in the Human Brain. *Cerebral Cortex*, *22*, 2943–2952.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annu Rev Neurosci*, *24*, 167–202.

Mischel, H. N., & Mischel, W. (1987). The development of children's knowledge of self-control strategies. In *Motivation, intention, and volition* (pp. 321–336). Springer.

Mischel, W., & Moore, B. (1980). The role of ideation in voluntary delay for symbolically presented rewards. *Cognitive Therapy and Research*, *4*, 211–221.

Munoz, D. P., & Everling, S. (2004). Look away: The anti-saccade task and the voluntary control of eye movement. *Nature Reviews Neuroscience*, *5*, 218.

Naragon-Gainey, K., McMahon, T. P., & Chacko, T. P. (2017). The structure of common emotion regulation strategies: A meta-analytic examination. *Psychological Bulletin*, *143*, 384.

Nee, D. E., Jonides, J., & Berman, M. G. (2007). Neural mechanisms of proactive interference-resolution. *Neuroimage*, *38*, 740–751.

Ochsner, K. N., & Feldman Barrett, L. (2001). A multiprocess perspective on the neuroscience of emotion. *Emotion: Current Issues and Future Directions*, 38–81.

Öhman, A., Flykt, A., & Esteves, F. (2001). Emotion drives attention: Detecting the snake in the grass. *Journal of Experimental Psychology: General*, *130*, 466.

Phelps, E. A., Ling, S., & Carrasco, M. (2006). Emotion facilitates perception and potentiates the perceptual benefits of attention. *Psychological Science*, *17*, 292–299.

Plato. (1952). *Phaedrus* (R. Hackforth, Trans.). Cambridge: Cambridge University Press.

Robinson, M. D., Schmeichel, B. J., & Inzlicht, M. (2010). A cognitive control perspective of self-control strength and its depletion. *Social and Personality Psychology Compass*, *4*, 189–200.

Scarantino, A. (2014). The motivational theory of emotions. *Moral Psychology and Human Agency*, 156–185.

Schachar, R. J., Tannock, R., & Logan, G. (1993). Inhibitory control, impulsiveness, and attention deficit hyperactivity disorder. *Clinical Psychology Review*, *13*, 721–739.

Scherer, K. (2001). *Appraisal Processes in Emotion: Theory, Methods, Research*. Oxford University Press.

Shenhav, A., Botvinick, M. M., & Cohen, J. D. (2013). The expected value of control: An integrative theory of anterior cingulate cortex function. *Neuron*, *79*, 217–240.

Shenhav, A., Cohen, J. D., & Botvinick, M. M. (2016). Dorsal anterior cingulate cortex and the value of control. *Nature Neuroscience*, *19*, 1286.

Shenhav, A., Musslick, S., Lieder, F., Kool, W., Griffiths, T. L., Cohen, J. D., & Botvinick, M. M. (2017). Toward a rational and mechanistic account of mental effort. *Annual Review of Neuroscience*, *40*, 99–124.

Skinner, M. D., & Aubin, H.-J. (2010). Craving's place in addiction theory: Contributions of the major models. *Neuroscience & Biobehavioral Reviews*, *34*, 606–623.

Smallwood, J., Fitzgerald, A., Miles, L. K., & Phillips, L. H. (2009). Shifting moods, wandering minds: Negative moods lead the mind to wander. *Emotion*, *9*, 271.

Smith, M. (2001). Responsibility and self-control. In *Relating to Responsibility: Essays for Tony Honoré on his Eightieth Birthday*. Hart Publishing.

Sripada, C. (2014). How is Willpower Possible? The Puzzle of Synchronic Self-Control and the Divided Mind. *Noûs*, *48*, 41–74.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*, 643.

Swinney, D. A., & Osterhout, L. (1990). Inference generation during auditory language comprehension. In *Psychology of Learning and Motivation* (Vol. 25, pp. 17–33). Elsevier.

Uleman, J. S., Adil Saribay, S., & Gonzalez, C. M. (2008). Spontaneous inferences, implicit impressions, and implicit theories. *Annu. Rev. Psychol.*, *59*, 329–360.

Wallace, R. J. (1999). Addiction as Defect of the Will: Some Philosophical Reflections. *Law and Philosophy*, *18*, 621–654.

Watson, G. (1977). Skepticism about Weakness of the Will. *Philosophical Review*, *86*, 316–339.

Wenzlaff, R. M., & Wegner, D. M. (2000). Thought suppression. *Annual Review of Psychology*, *51*, 59–91.

Willcutt, E. G., Doyle, A. E., Nigg, J. T., Faraone, S. V., & Pennington, B. F. (2005). Validity of the executive function theory of attention-deficit/hyperactivity disorder: A meta-analytic review. *Biological Psychiatry*, *57*, 1336–1346.

Figure Captions

Figure 1: Core Architecture of Emotion-Type States. *Emotion-type states are a diverse collection of states that includes emotions, drives, impulses, cravings, and pains.* ***.*** *They are elicited by non-deliberative appraisal processes. Once elicited, they produce biases across multiple psychological mechanisms. As a consequence of these biases, these mechanisms produce temporally-extended streams of state-congruent response pulses. These response pulses, in turn, strongly bias the selection of actions that are congruent with the emotion-type state. It is argued in the main text that the elements in green boxes can be targets of control while those in red boxes cannot.*