RUNNING HEAD: NEUROSCIENCE OF SELF-CONTROL

The neuroscience of self-control

Elliot T. Berkman

Department of Psychology, University of Oregon

Center for Translational Neuroscience, University of Oregon

Abstract word count: 85

Main text word count: 5937

Full chapter word count: 7297 (of 7500)

References: 50 (of 50)

Figures: 5

Address correspondence to:

Elliot T. Berkman
Department of Psychology
1227 University of Oregon
Eugene, OR 97403-1227
berkman@uoregon.edu

In press in D. de Ridder, M. Adriaanse, & K. Fujita (Eds.), *Handbook of Self-Control in Health and Wellbeing*. Abingdon-on-Thames: Routledge.

Abstract

This chapter describes why and how neuroscience methods can be useful for self-control theory and research. The author contrasts two models of self-control, the opposition and valuation model, with an emphasis on how each characterizes the mechanisms self-control. Next, the author reviews recent neuroscience relevant to each model derived from functional magnetic resonance imaging studies. The author closes by offering an integrated account of how self-control operates at a neurocognitive level, and suggesting ways that self-control might be improved in light of the neurally-informed model.

It seems that no literature within psychology is complete these days without some data on the neural correlates of the topic. This is true of self-control, which is certainly above the mean and might even be an outlier in terms of the exceptionally large quantity of relevant neuroscientific data. The purpose of this chapter is to provide a useful framework for thinking about those data rather than to provide a comprehensive review of them. Another goal is to critically evaluate the possible contributions of neuroscience to the study of self-control.

I begin by articulating several reasons why knowledge of brain function can improve psychological models of self-regulation. I then describe classes of two models of self-control that have guided neuroscientific study. Next, I explain how the neuroscientific data from the models are largely convergent and outline a few points of difference that need to be reconciled. I close by describing future directions for self-control research where neuroimaging could be particularly impactful.

Why should self-control researchers study the brain?

I do not take it for granted that obtaining a map that links brain regions to mental processes will necessarily be useful for understanding self-control. There are many reasons I do not make this assumption, but two are especially pertinent here. First, the map between a given brain region and a given mental process is many-to-many. The neural regions recruited during self-control are involved in many other processes, so merely observing that one or more of those regions was active during a task does not necessarily imply that self-control occurred. This so-called “reverse inference” problem limits the ability of researchers to infer mental process from brain data alone (Poldrack, 2006). Second, the number of tools that exist to alter brain function directly and with some degree of specificity is very small. So, even if research were to identify a brain region or network that is causally involved in effective self-control (e.g., self-control cannot be performed if that region or network is lesioned, and self-control only requires that region or network), interventions would be unlikely to be able to target it. I refer to this as the "so what?" problem because it limits the significance of obtaining even high-quality inferential knowledge of how the brain executes self-control.

However, there are also reasons to be hopeful. The root of the reverse inference problem is that the mapping between mental processes and neural activations is complex, but not ultimately unknowable. Accurate reverse inference is possible given sufficient information about the base rates of activation and the likelihood of a task invoking a particular process (Poldrack, Kittur, Kalar, Miller, Seppa, Gil, et al., 2011). If activation in Region X is rare across all cognitive neuroscience studies but common in studies that elicit Process A, then there is a reasonable chance that Process A is involved in a new task if activation is observed in Region X. Indeed, a central purpose of NeuroSynth, a software platform of large-scale automated meta-analysis of neuroimaging data, is to uncover region-specific base rates of activation to enable valid reverse inference (Yarkoni, Poldrack, Nichols, Van Essen, & Wager, 2011). This tool, as well as the broader movement toward Bayesian approaches that leverage prior information about a phenomenon to refine scientific knowledge, are only just beginning. The ability of scientists to infer mental processes from brain activation will grow rapidly as more knowledge accumulates and the tools available to take advantage of that knowledge are developed.

The "so what" problem relating to the difficulty of directly altering brain function is also offset for two reasons. First, in extreme cases, it is possible to alter the brain through direct surgical or pharmacological manipulation. Deep brain stimulation of the subgenual anterior cingulate cortex, for example, can be effective against treatment-resistant forms of depression (Mayberg, Lozano, Voon, McNeely, Seminowicz, Hamani, et al., 2005), and certain classes of drugs can effectively treat substance use by acting on receptors that otherwise would bind to the abused substance (Le Foll, Ciano, Panlilio, Goldberg, & Ciccocioppo, 2013). Emerging neurostimulation methods such as transcranial direct current stimulation (tDCS) also enable researchers to manipulate brain activity in a less invasive way, and these methods have been shown to be effective in altering brain function in specific neural areas implicated in disorder (Nitsche, Boggio, Fregni, & Pascual-Leone, 2009).

Second, it is also possible to alter brain function through indirect routes. A novel and innovative class of psychosocial “brain-training” interventions are beginning to emerge that can target key systems, such as the regions of the frontrostriatal motor planning and implementation network that are involved in inhibitory control (Berkman, Kahn, & Merchant, 2014; see Bryck & Fisher, 2012, for a summary). These interventions use behavioral or psychosocial means, such as narrowly focused neurocognitive tasks, to engage and thereby alter the function of specific neural systems. Brain-training protocols are grounded in basic cognitive, affective, and social neuroscience research that identifies procedures to elicit activity in the targeted networks. Discovery of new protocols is somewhat haphazard at this preliminary stage in the field’s development because a coherent framework for intervention development is lacking: Given a specific neural system, what is the procedure for creating an intervention that might alter it? A logical starting point is training based on associative learning or classical conditioning. Examples of recent successes based on these approaches include cognitive training to target proactive control in the lateral prefrontal cortex (Berkman et al., 2014) and attention bias modification to target amygdala activation in anxiety (Britton, Suway, Clementi, Fox, Pine, & Bar-Haim, 2015). Progress in this area will be made as researchers build more sophisticated frameworks for understanding how the brain responds to specific forms of training by incorporating advances in neuroscientific knowledge (see Beauchamp, Kahn, & Berkman, under review, for more discussion of this topic).

Regardless of whether the nature of the interventions, they will depend critically on how researchers conceptualize the nature of self-control. We now turn to describing two models of self-control.

How does self-control work?

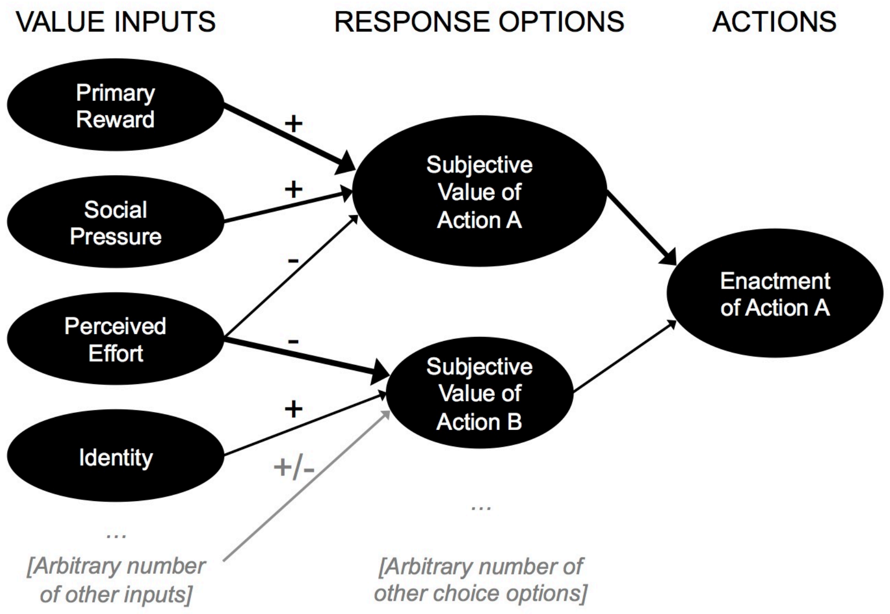
Several families of definitions for self-control have emerged that overlap considerably but make diverging predictions about the mechanisms of self-control. Traditional models characterize self-control as an integration of—and sometimes competition between—bottom-up, "hot" processes such as reward responsivity against top-down, "cold" processes such as inhibitory control (Baumeister & Heatherton, 1996; Metcalfe & Mischel, 1999). Other models focus on the conflict between long-term goals versus short-term goals or temptations (Carver & Scheier, 1998), even if those models are more agnostic about the specific processes that contribute to the resolution of that conflict. A hallmark of this family of models is the "opposition assumption" (Kahn & Berkman, under review) that various bottom-up processes (e.g., reward, temptation, impulse) oppose and are opposed by top-down processes (e.g., cognitive control). The oppositional or inhibitory nature of these two classes of processes is captured by a see-saw metaphor (Heatherton & Wagner, 2011), whereby self-control failure is characterized by excessive activation of the bottom-up processes, insufficient activation of the top-down processes, or some combination thereof. It is important to note that not all dual-process models of self-control limit the interaction between the processes to inhibition (e.g., "hot" signals can bias the processing of the "cold" system without turning it off, Metcalf & Mischel, 1999), but all of these models take the general stance that bottom-up processes impede self-control and top-down processes promote it.

Attempts to identify the neural systems involved in self-control followed the opposition assumption, and have been successful. Broadly speaking, activation of limbic system regions is associated with temptation and indulgence, whereas activation of lateral prefrontal regions is associated with self-control engagement and success. When activation in both sets of is measured, there is often an inverse relationship between the two, and that the degree of inverse association is linked to self-control success (see Buhle, Silvers, Wager, Lopez, Onyemekwu, Kober, Weber, et al., 2014, for a meta-analysis). For example, emotion regulation using cognitive reappraisal of upsetting negative images (Ochsner & Gross, 2008) and appetitive food cues (Giuliani, Mann, Tomiyama, & Berkman, 2014) increases activity in dorsolateral and ventrolateral prefrontal cortices and decreases activity in amygdala and ventral striatum, respectively, and increases the inverse coupling between the two systems (Banks, Eddy, Angstad, Natha, & Phan, 2007). When measured, activity in lateral prefrontal regions and its connections with subcortical regions also tracks with self-rated control success (Ochsner, Ray, Cooper, Robertson, Chopra, Gabrieli, & Gross, 2004). Similar patterns conforming to the opposition assumption have been observed among cigarette smokers controlling urges to smoke (Kober, Mende-Siedlecki, Kross, Weber, Mischel, Hart, et al., 2010), cocaine abusers controlling drug craving (Volkow, Fowler, Wang, Telang, Logan, Jayne, et al., 2010) and in clinical populations (e.g., Goldin, Manber-Ball, Werner, Heimberg, & Gross, 2009). At a first pass, the broad pattern of lateral prefrontal regions down-regulating subcortical ones appears to hold for both appetitive (e.g., cravings) and aversive (e.g., distress) responses.

However, the opposition assumption is based on a metaphorical model of self-control (e.g., a see-saw or a horse-and-rider) and was never intended to provide concrete predictions about how the brain actually solves self-control dilemmas. The distinction between "hot" and "cold" processes, for example, is certainly useful in understanding the clear phenomenological difference between temptation and inhibition—they *feel* different—but does not mean that these two processes are necessarily distinct from each other or are each unitary at the neural level. Despite some initial success of oppositional models, a recent wave of studies that investigate the neuroscience of self-control with more sophisticated methods and ecologically valid stimuli has revealed there to be far more complexity than can be encompassed by a simple top-down versus bottom-up model. For example, increased activation in the lateral prefrontal cortex is linked with self-control success in some cases (e.g., Demos, Kelley, & Heatherton, 2011) and failure in others (David, Munafo, Johansen-Berg, Smith, Rogers, Matthews, et al., 2005). Also, the specific region or set of regions involved in self-control varies over a large swath of cortex from study to study (Berkman & Lieberman, 2009). The authors of a prominent review noted that "the field has struggled to coalesce around a unified view of the control mechanisms that support self-regulation" (Kelley, Wagner, & Heatherton, 2015, pp. 390). Although there is some evidence for an oppositional model of self-control at the level of the brain, the lack of consistency suggests that additional processes might be at play. A model that aims to explain the existing evidence needs to account for conditions when top-down and bottom-up processes are aligned as well as when they are opposed.

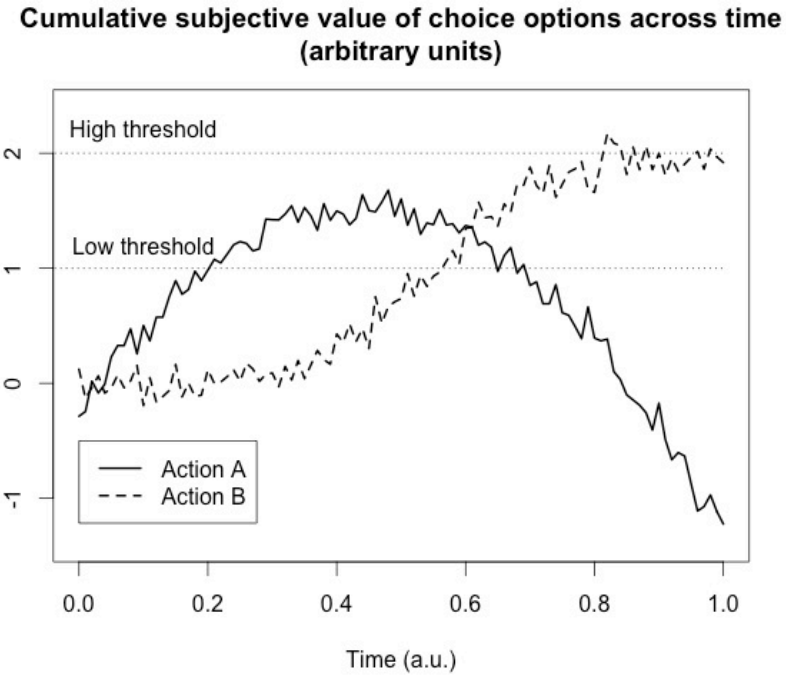
For the field to move toward such a unified view, more flexible and comprehensive models will be needed to make sense of the varied set of separable processes subserved by the prefrontal cortex. These new models might be viewed by some as competing with existing models, but that need not be the case. The existing models are adequate in some conditions but not others. What the models need, therefore, is to be updated to extend their range of applicability. Often this extension comes in the form of mechanistic elaboration. A similar progression has occurred in research on executive functions within cognitive neuroscience, where it has been eloquently articulated that the psychological construct of "top-down control" can still be meaningful even though it refers to a heterogeneous mix of functions and processes at the neural level (Miller & Cohen, 2001). Along that line, a useful example has been provided by Kotabe and Hofmann (2015), who described an integrative self-control theory. In their theory, control and desire conflict in a broad sense—ultimately, self-control is defined by a conflict between two possible options. But the specific neurocognitive processes that resolve the conflict and decide which of the two behaviors is enacted are many (seven, in this model) and varied (some "hot" processes contribute to control, such as motivation), and the activation of these processes is context dependent. For instance, successful self-control can be caused either by very high control motivation or by fast conflict detection coupled with strong control effort. This model nicely illustrates how it is possible to preserve the phenomenology of an opposition between "hot" and “cold” while also offering a mechanistic account of the various ways that such an opposition plays out. Importantly, this model allows for the possibility of self-control success in the absence of top-down control over bottom-up processes, and of self-control failure in the absence of bottom-up processes. The model can explain all of these possibilities because it relaxes key assumptions of earlier models about the number of processes involved and the specific direction of their interactions.

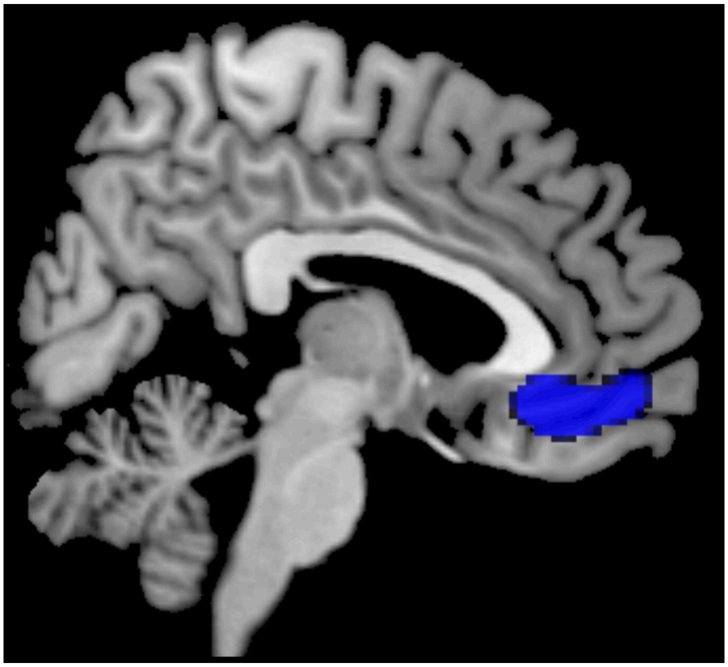
A parallel conversation took place within neuroscience as simple “prefrontal versus subcortical” models of cognition evolved to be more detailed and comprehensive. Fortuitously, such a neurocognitive model has been developed in the arena of choice and decision-making, and it has now been applied systematically to study self-control (e.g., Rangel & Hare, 2010). This model suggests that self-control dilemmas can be solved by the same neural machinery deployed for other decisions, broadly construed. On this view, self-control is a special case of decision-making, whereby one (or more) response options promote(s) a long-term goal and one (or more) promote(s) short-term goals, temptations, or otherwise do not promote the long-term? goal. Self-control success is defined as the act of choosing an option that promotes a long-term goal. Like the Kotabe and Hofmann (2015) model, this model decomposes the deceptively simple problem of picking one of two alternatives by allowing for the possibility that a variety of underlying mechanisms might pull for or against either option, or not, depending on conditions. The magnitude of the “pull for” is reward value, and the “pull against” is cost. Critically, each choice option can potentially have many sources of value and cost represented throughout the brain (Rangel, Camerer, & Montague, 2008). These heterogeneous sources of value and cost are translated into a “common currency” and integrated in the ventromedial prefrontal cortex (vmPFC), which serves as a central locus that tracks the cumulative subjective value of options in a decision (Levy & Glimcher, 2011).


VALUE INPUTS
RESPONSE OPTIONS
ACTIONS
Primary Reward
Social Pressure
Perceived Effort
Identity
...
[Arbitrary number of other inputs]
+
+
-
-
+
+/-
Subjective Value of Action A
Subjective Value of Action B
...
[Arbitrary number of other choice options]
Enactment of Action A

As applied to self-control, the valuation model predicts that each option in a self-control conflict accumulates subjective value based on an arbitrary number of value inputs (Figure 1). In psychology studies, there are usually only two choice options, but in the real world there can be many more (Figure 1 shows only two options for simplicity). Similarly, each option can have many value inputs. The value inputs can fluctuate dynamically depending on the organism's changing needs, available resources, and attentional focus. For example, the relative value of choice options can change as a function of which options are included in the choice set (Tversky & Simonson, 1993); they can also change depending on levels of scarcity (Shah, Shafir, & Mullainathan, 2015). It follows that which options are noticed and evaluated, and which are ignored or unseen, is a major factor in determining self-control outcomes. Additionally, the reference point against which values are assigned is not an absolute value, but rather a relative value that can change depending on psychological factors such as framing relative to a reference value (Kahneman & Tversky, 1979). Therefore, the outcome of a self-control dilemma is a product not only of the value inputs, but also of the context, the choice set, and the reference point, all of which can change from moment to moment.

The total value of each choice option accumulates across time as various properties of the options are considered. Eventually, the option with the greatest value is enacted when a threshold is reached or time runs out to make a decision. This process can be captured with a stochastic evidence accumulator model, which compiles noisy data until a threshold is reached or a decision must be made (see Figure 2; Smith & Ratcliff, 2004). Value is calculated in an ongoing, cumulative manner in the moments leading up to a decision, and thus fluctuates over time as various input sources are integrated dynamically. A consequence is that the option with the greatest subjective value can change over time as relative values change. This provides a simple explanation for the often-impulsive nature of speeded choices: the value of long-term goals takes longer to accumulate than that of immediately pleasurable experiences (Sullivan, Hutcherson, Harris, & Rangel, 2015).





Neuroeconomics research overwhelmingly implicates regions in the mesolimbic dopaminergic system, primarily the vmPFC and also the orbitofrontal cortex (OFC) and ventral striatum (vS), in the integration of subjective value (Figure 3). Indeed, the vmPFC and the OFC are considered the same area by some (e.g., Pearson, Watson, & Platt, 2014). Consistent with the common currency idea, this research suggests that the vmPFC is involved in the computation of subjective value of both appetitive and aversive stimuli (Bartra, McGuire, & Kable, 2013; Tom, Fox, Trepel, & Poldrack, 2007). In a series of studies, Rangel and colleagues have found that the vmPFC integrates information across a range of properties about a stimulus to produce a final value signal that integrates stimulus properties, active goals, costs, and other types of choice-relevant information (Rangel & Hare, 2010). Specifically, the vmPFC receives inputs from regions associated with bottom-up processes (e.g., vS) and from regions associated with top-down processes (e.g., dlPFC). Not only does vmPFC gather various value sources, but activity in the vmPFC also tracks the subjective value of a range of stimulus types (Padoa-Schioppa & Assad, 2006). For example, vmPFC activity predicts choice regardless of whether the stimuli under evaluation depict food or money (Levy & Glimcher, 2011). A related study found that activity in vmPFC scales with the subjective value of a monetary gain both for oneself and for another person (Zaki, Lopez, & Mitchell, 2014). These findings converge in identifying the vmPFC as playing a central role in the integration of subjective value from both “hot” and “cold” inputs.

The presumptive purpose of this vmPFC “unified valuation system” that integrates across disparate outcomes is to facilitate choice among them (Levy & Glimcher, 2011). For example, the vmPFC value signal predicts decisions regardless of whether they appear to be driven by processes related to impulsivity or restraint (e.g., keeping money vs. giving it to charity, or eating unhealthy vs. healthy foods; Hare, Camerer, Knoepfle, O’Doherty, & Rangel, 2010; Hare, Malmaud, & Rangel, 2011a). In another study, participants separately rated the tastiness and healthiness of a series of food stimuli, and then made choices about whether or not to eat each food (with one choice randomly selected at the conclusion of the study and given to the participant to eat). Activity in vmPFC predicted the participants' subsequent choices regardless of whether the choice on a given trial was driven by health or taste concerns (Hare, Camerer, & Rangel, 2009). The vmPFC thus appears to be a point of accumulation for value-related information that contributes to subsequent choice.

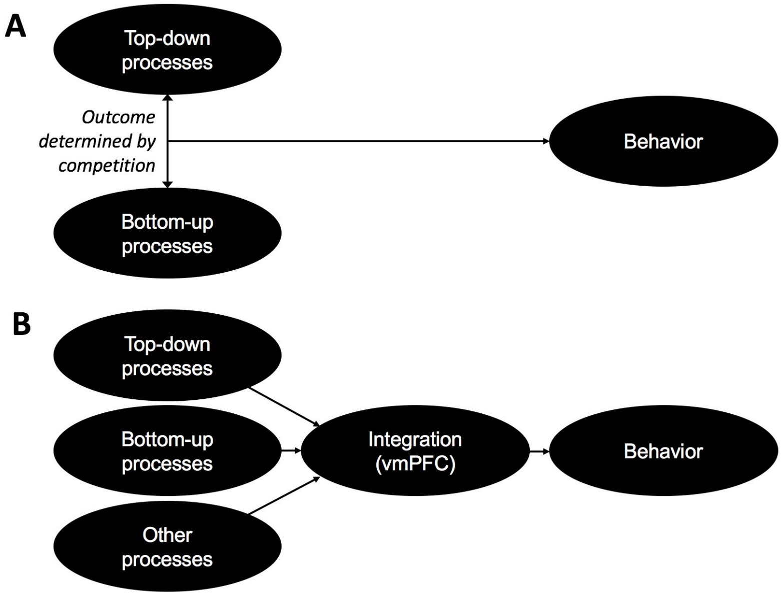
The valuation model predicts that the value calculation integrates different kinds of inputs in a flexible way depending on the context. Consistent with this idea, vmPFC receives inputs from different brain regions depending on the contextual cues and available response options. For example, the dorsolateral prefrontal cortex (dlPFC) increases its functional connectivity with the vmPFC when higher-order goals such as health concerns or social factors are made salient during food choice; otherwise, vmPFC mostly receives input from regions encoding primary reward associated with tastiness (Hare et al., 2010; 2011a; Hutcherson, Plassman, Gross, & Rangel, 2012). There is also evidence that the value of response options are reflected in the vmPFC before specific action plans are selected (Wunderlich, Rangel, & O'Doherty, 2010), and that value signals provide input to downstream brain regions that are responsible for selecting and implementing motor plans (Hare, Schultz, Camerer, O'Doherty, & Rangel, 2011b). And, like subjective value during choice, activity in the vmPFC in the moments before a decision also fits a stochastic evidence accumulator model (De Martino, Fleming, Garrett, & Dolan, 2013). Taken together, then, the emerging view from the neuroeconomics literature is that the vmPFC represents a point of convergence for a variety of input signals that are relevant to the decision at hand, and its activation reflects a dynamic value integration process that subsequently biases behavior toward high-valued actions.

Above, we have reviewed evidence showing that activity in vmPFC and related regions (a) represents the subjective value of a variety of stimuli, (b) temporally precedes and predicts choice, and (c) receives input from other brain regions depending on the context and choice options. This strongly implicates the vmPFC as the neuroanatomical locus of an integration of heterogeneous inputs into a common neural currency of subjective value that influences decisions.

Comparing and synthesizing the models

The valuation and opposition models of self-control models appear at first glance to be at odds. Valuation models treat self-control as a choice, whereas opposition models characterize self-control as a battle between the strength of qualitatively different processes; valuation models focus on integration in the vmPFC, whereas opposition models posit that connectivity between lateral prefrontal and subcortical regions is key. Upon closer inspection, however, these differences are revealed to be matters of terminology and level of specificity rather than substance. At their core, the models converge far more than they diverge. For example, most opposition models do not specify exactly how the interaction between top-down and bottom-up processes play out. But if they allow for the possibility of a third, intermediary process that adjudicates between the two, then the difference between opposition and valuation models disappears. Here, I describe how the models might be harmonized.

Self-control involves a conflict between two or more behavioral options. Often, in the real world and in the laboratory, one of the options is related to a "hot" process (e.g., a desire to eat a delicious-looking piece of cake) and another is related to a "cold" process (e.g., the abstract representation of the goal to eat less sugar). A key distinction between the models in question is in how they predict that conflict gets resolved. Some opposition models suggest that the two processes interact directly with one another (e.g., through inhibition), but most are silent on the exact process. Valuation models provide a more detailed mechanistic account by suggesting that each input process contributes separately to a unified value calculation (Figure 4). These two scenarios are not all that different from each other. In both, some top-down process propels action toward one behavior and some bottom-up process propels action toward a different one. Thus, in the simple tasks that we use in laboratory neuroimaging studies of self-control that deliberately evoke only two processes (e.g., cognitive reappraisal of negative emotion), the models are nearly indistinguishable.



The simple addition of an intermediate integration step enables valuation models to explain a far greater range of data than strict opposition models. For example, in opposition models, it is unclear how both decreased *and* increased lateral prefrontal activation (presumably an index of top-down processing) can be associated with successful self-control. This tension can be resolved in two ways. First, this inconsistency can be addressed by decoupling top-down and bottom-up processes, which allows for the possibility that self-control can be achieved without top-down processing—as long as bottom-up processing is reduced by some other means. Second, valuation models explicitly allow for inputs from an arbitrary number of inputs to a decision instead of just two, and are less rigid about the top-down or bottom-up nature of those processes. This is possible in opposition models, too, but they do not specify how third systems influence the central dynamic between top-down and bottom-up processes. For example, social factors such as peer-influence or perspective-taking frequently contribute to self-control decisions. In the real world, sometimes all three of these processes are active simultaneously: a smoker might be tempted to have a cigarette and recruit inhibitory control while also considering what his peers would think of his lapse. Opposition models place social factors into the "top-down" or "bottom-up" category depending on whether it promotes or prevents self-control, which confounds the qualitative nature of the process with the behavior it promotes. In contrast, valuation models account for these kinds of situations by positing that each source of value contributes independently to a unified value calculation. This difference between the models can be resolved by specifying in opposition models how and when third processes such as social influence affect self-control.

The two models present slightly different characterizations of what functions are subserved by lateral prefrontal regions. As with the other distinctions between the models, this is subtle but has major implications for neuroscientific investigations into self-control. As alluded to previously, opposition models generally confound the top-down or bottom-up nature of the process with the goal-promoting or goal-preventing status of the behavior they facilitate. Top-down is taken a priori to mean goal-promoting, and bottom-up to mean goal-preventing. This equation is reasonable in laboratory studies, where experimenters deliberately construct the situation in such a way. However, there is no necessary equivalence between the two at the level of the brain. Top-down processes instantiated in the prefrontal cortex can be goal-preventing, such as when dieters elaborately plan indulgent meals and concoct rationalizations for this behavior (e.g., Thanksgiving), or when goal-promoting behavior in smokers is incentivized with other forms of reward (e.g., contingency management treatment). Understanding self-control at the behavioral level, therefore, is less about the nature of the processes or their associated neural activations and more about the magnitude and direction of their contribution to the decision of which behavior to enact. Valuation models focus squarely on this issue by characterizing the inputs in terms of values (and costs), regardless of whether those inputs derive from top-down or bottom-up processes or where in the brain they originate. In this case, one top-down goal competed with another, causing self-control success at one goal (hosting a fun dinner) at the cost of another (dieting). Opposition models can be expanded to account for this by characterizing the opposition as between goal-promoting and goal-preventing behaviors rather than as between top-down and bottom-up processes. The opposition assumption is still useful in the (frequent) case that top-down and bottom-up processes are aligned with goal-promoting goal-preventing behaviors, respectively, but that alignment does not always hold.

The subtle shift in the characterization of prefrontal cortical activation from top-down goal-promotion to simple goal value representation (following Miller & Cohen, 2001) highlights the fact that humans operate on multiple goals in parallel. The goal or goals that are active shift across time, compete with one another at times, and are not always known to the experimenter. In teenage smokers, for example, the goals of being healthy, on one hand, and being accepted into a peer group, on the other, conflict with each other if the teens believe that smoking will earn them acceptance with their peers. These goals are both represented in the prefrontal cortex, though perhaps in different locations (e.g., health goals in dorsolateral prefrontal cortex, social goals in temporoparietal regions, Hare et al., 2010; 2011a). Valuation models provide a parsimonious explanation of how these goal-goal conflicts play out: each contributes some degree of value to a unified calculation, and the highest-valued goal is enacted. Opposition models focus on the conflict between the hedonic desire to smoke and the top-down resources to control that desire (which presumably depends on the strength of the health goal). In this way, opposition models artificially narrow the range of processes considered in understanding self-control and its neural underpinnings, and are unnecessarily burdened with normative definitions of what goal success and failure mean to an individual at a given time. This can be resolved by relaxing the assumptions about what sorts of processes promote or prevent self-control, and how many there are.

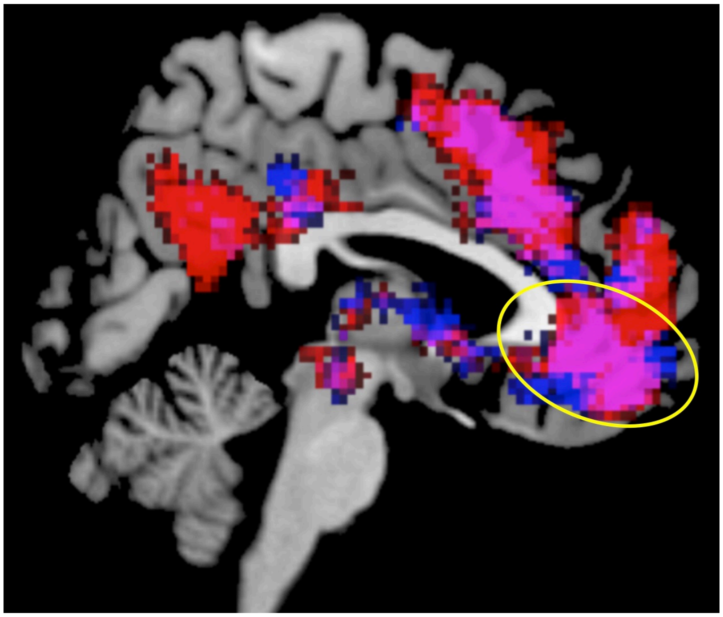
One final distinction between the models deals with the heterogeneity of the processes involved in various self-control efforts and its neuroanatomical implications. For example, top-down processes relevant to self-control include working memory, inhibitory control, and linguistic reconstrual, and bottom-up processes include hedonic temptation and negative affective reactivity such as fear or disgust. It is uncontroversial to claim that the activation of each of these processes recruits a different set of neural regions, even if the activations likely overlap to some extent. For the purposes here, however, opposition models require that each top-down process has connectivity with each bottom-up process, either direct or indirect. For example, linguistic reconstrual has been shown to decrease both negative responses to fear-inducing stimuli (Ochsner et al., 2004) and positive responses to appetitive stimuli (Kober et al., 201), represented in the amygdala and vS, respectively. Does this happen through direct or indirect connections? Suppose there are 4 top-down regions and 4 bottom-up regions involved in self-control; for each top-down region to be linked with each bottom-up region requires $4^2 = 16$ pathways. In general, the total number of pathways required for complete top-down to bottom-up connectivity is the square of the number of specific locations on each side. In contrast, if the conflict is resolved through an intermediate region (e.g., vmPFC) to which all top-down and bottom-up regions have connectivity, then the total number of links is equal to two times the number of regions on each side, or 4*2 = 8, in the example. Thus, any model where oppositional regions interact through an adjudicator is more anatomically efficient than models where such regions interact directly with each other. This is explicitly posited in valuation models and potentially consistent with opposition models but often unaddressed.

This section highlighted some of the major similarities and key differences between opposition models and valuation models. The differences between them are largely subtle, and more easily detected using different paradigms than the ones deployed in standard laboratory tasks of self-control, which deliberately confound top-down/bottom-up with goal success/failure (Kahn &

Berkman, under review). These kinds of tasks have been so popular in the field because they are representative of how people think of self-control. Metaphors for the self-control struggle such as the horse and rider or the devil on one shoulder and the angel on the other date back to the Greeks. Perhaps most self-control in the real world follows that pattern. Nonetheless, a few subtle shifts in their focus and assumptions allow valuation models to account for a far wider range of possibilities (e.g., third processes, competing top-down goals) than opposition models, and do so with greater anatomical and computational efficiency. In the final section, we consider important next steps for the neuroscientific study of self-control.

Future directions

One of the most substantive differences between opposition and valuation models of self-control is in their prescriptions for improving self-control. Opposition models suggest that strengthening top-down control will improve self-control. Despite the intuitive nature of this prediction, it is increasingly apparent that training in effortful, top-down control does not improve self-control beyond the focal (i.e., trained) task (Berkman, in press). Other options suggested by opposition models are to reduce the strength of the temptation through learning processes or by avoiding the tempting stimulus entirely (see Heatherton & Wagner, 2011, for a review of the ways bottom-up processes can foster self-control failure). These techniques surely work, but what advice does psychology research offer to would-be self-controllers who want to improve their abilities when faced with a temptation? The valuation model indicates that any intervention that increases the subjective value of a goal-promoting behavior would facilitate self-control. This orients researchers to a host of value-based interventions that are generally outside the scope of opposition models such as incentives (e.g., contingency management), social norms manipulations, and identity-based programs. Ongoing research in my lab focuses on the latter given the close anatomical relationship between the vmPFC, which is involved in valuation, and highly overlapping aspects of the medial prefrontal cortex, which is involved in self and identity (Figure 5). Motivation and value are difficult to measure through self-report but the vmPFC has a well-established role in valuation, suggesting an important role for neuroimaging in developing and validating these interventions.



There are also several neuroanatomical challenges that remain to be addressed. The most prominent is to build a “neuromotivational” map of the brain that indicates what kinds of value inputs are represented where. The map is not entirely blank at this point—we have a rough sense that hedonic and primary rewards are represented in mesolimbic dopamine structures such as the vS, abstract goal value is represented in lateral prefrontal cortices, and social reward value is represented in parts of the mentalizing network—but we have much to learn. More refined knowledge of how the brain represents different forms of value will enable researchers to identify ways to leverage those regions for improved self-control, and to identify interventions that might be more or less effective for a given person given his or her brain function and structure. Also, the degree of bidirectionality between these input regions (e.g., lateral prefrontal cortex, amygdala) and the vmPFC still needs to be established. These connections appear to be anatomically present (Ongür & Price, 2000), but whether they come online during self-control decisions is unknown. Valuation models would have difficulty accounting for observed inverse correlations between lateral prefrontal and subcortical regions (e.g., Banks et al., 2007) in the absence of a feedback process from vmPFC to the input regions. Such a process could also explain balance and consistency effects in psychology by providing a mechanism whereby ultimate preferences and choices feed back to influence the original evaluation of a stimulus or action.

Finally, future research should strive to incorporate goal-relevant behavior into neuroimaging studies of self-control. It is notable that nearly all of the neuroimaging studies in support of opposition models are based in paradigms where the outcome is entirely intrapersonal (e.g., cognitive reappraisal of negative emotions or craving), whereas much of the neuroimaging support for valuation models deploy behavioral economics paradigms which require participants to play for real stakes with real money. These paradigms cleverly address the necessity for multiple trials in functional neuroimaging research by telling participants that one trial will be randomly selected to enactment (e.g., Hare et al., 2009), so the best strategy for participants is to engage with every trial as though their decision on that trial would lead to an actual purchase. Further, participants know these exchanges are real because behavioral economics studies never use deception. It is possible, therefore, that many of the observed differences (e.g., the presence or absence of vmPFC activation during self-control as noted in Kelley et al., 2015) may be due to the availability (or lack) of an actual behavioral output. Ecological validity is a perennial challenge in neuroimaging research (c.f., Berkman & Lieberman, 2009), but one that must be given high priority when the phenomenon at hand is so centrally behavioral, as is the case with self-control.

Conclusion

I described the challenges of studying self-control using neuroimaging methods as well as some of the advantages. I presented two major classes of models of self-control that focus on its constituent mechanisms and their neural underpinnings. The models are largely similar though differ in whether they view top-down and bottom-up processes as necessarily oppositional in the context of self-control. Studies motivated by both models have contributed substantial knowledge about the neuroscience of self-control. The field is now moving toward an integrative arena where this knowledge can fruitfully be synthesized into a broader model that, ideally, connects to ideas from other areas of psychology and in allied fields. A major future direction for self-control research using neuroscience is to find ways to improve self-control based on insights that would not have been apparent without knowledge about basic brain systems.

References

Banks, S., Eddy, K., Angstadt, M., Nathan, P., & Phan, K. (2007). Amygdala-frontal connectivity during emotion regulation. *Social Cognitive and Affective Neuroscience*, *2*, 303–312.

Bartra, O., McGuire, J. T., & Kable, J. W. (2013). The valuation system: A coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *NeuroImage*, *76*, 412–427.

Baumeister, R. F., & Heatherton, T. F. (1996). Self-regulation failure: An overview. *Psychological Inquiry*, *7*, 1–15.

Beauchamp, K. G., Kahn, L. E., & Berkman, E. T. (under review). Does inhibitory control training transfer? Behavioral and neural effects on an untrained emotion regulation task.

Berkman, E.T. (in press). Self-regulation training. In K. D. Vohs & R. F. Baumeister (Eds.), *Handbook of Self-Regulation: Research, Theory and Applications (3rd Edition)*. New York: Guilford.

Berkman, E. T., & Lieberman, M. D. (2009). Using neuroscience to broaden emotion regulation: Theoretical and methodological considerations. *Social and Personality Psychology Compass*, *3*, 475–493.

Berkman, E. T., Kahn, L. E., & Merchant, J. S. (2014). Training-induced changes in inhibitory control network activity. *The Journal of Neuroscience*, *34*, 149–157.

Britton, J. C., Suway, J. G., Clementi, M. A., Fox, N. A., Pine, D. S., & Bar-Haim, Y. (2014). Neural changes with attention bias modification for anxiety: a randomized trial. *Social Cognitive and Affective Neuroscience*, *10*, 913–920.

Bryck, R. L., & Fisher, P. A. (2012). Training the brain: Practical applications of neural plasticity from the intersection of cognitive neuroscience, developmental psychology, and prevention science. *The American Psychologist*, *67*, 87–100.

Buhle, J. T., Silvers, J. A., Wager, T. D., Lopez, R., Onyemekwu, C., Kober, H., et al. (2014). Cognitive reappraisal of emotion: a meta-analysis of human neuroimaging studies. *Cerebral Cortex*, *24*, 2981–2990.

Carver, C., & Scheier, M. (1998). On the self-regulation of behavior. New York, NY: Cambridge Univ Press.

David, S. P., Munafò, M. R., Johansen-Berg, H., Smith, S. M., Rogers, R. D., Matthews, P. M., & Walton, R. T. (2005). Ventral striatum/nucleus accumbens activation to smoking-related pictorial cues in smokers and nonsmokers: A functional magnetic resonance imaging study. *Biological Psychiatry*, *58*, 488–494.

De Martino, B., Fleming, S. M., Garrett, N., & Dolan, R. J. (2013). Confidence in value-based choice. *Nature Neuroscience*, *16*, 105–110.

Demos, K. E., Kelley, W. M., & Heatherton, T. F. (2011). Dietary restraint violations influence reward responses in nucleus accumbens and amygdala. *Journal of Cognitive Neuroscience*, *23*, 1952–1963.

Giuliani, N. R., Mann, T., Tomiyama, A. J., & Berkman, E. T. (2014). Neural systems underlying the reappraisal of personally craved foods. *Journal of Cognitive Neuroscience*, *26*, 1390–1402.

Goldin, P. R., Manber-Ball, T., Werner, K., Heimberg, R., & Gross, J. J. (2009). Neural mechanisms of cognitive reappraisal of negative self-beliefs in social anxiety disorder. *Biological Psychiatry*, *66*, 1091–1099.

Hare, T. A., Camerer, C. F., & Rangel, A. (2009). Self-control in decision-making involves modulation of the vmPFC valuation system. *Science*, *324*, 646–648.

Hare, T. A., Camerer, C. F., Knoepfle, D. T., O'Doherty, J. P., & Rangel, A. (2010). Value computations in ventral medial prefrontal cortex during charitable decision making incorporate input from regions involved in social cognition. *The Journal of Neuroscience*, *30*, 583–590.

Hare, T. A., Malmaud, J., & Rangel, A. (2011a). Focusing attention on the health aspects of foods changes value signals in vmPFC and improves dietary choice. *Journal of Neuroscience*, *31*, 11077–11087.

Hare, T. A., Schultz, W., Camerer, C. F., O'Doherty, J. P., & Rangel, A. (2011b). Transformation of stimulus value signals into motor commands during simple choice. *Proceedings of the National Academy of Sciences*, *108*, 18120–18125.

Heatherton, T. F., & Wagner, D. D. (2011). Cognitive neuroscience of self-regulation failure. *Trends in Cognitive Sciences*, *15*, 132–139.

Hutcherson, C. A., Plassmann, H., Gross, J. J., & Rangel, A. (2012). Cognitive regulation during decision making shifts behavioral control between ventromedial and dorsolateral prefrontal value systems. *The Journal of Neuroscience*, *32*, 13543–13554.

Kahn, L. E., & Berkman, E. T. (under review). Incentivizing inhibitory control with reward: When hot and cold processes are not opposed.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*, 263–292.

Kelley, W. M., Wagner, D. D., & Heatherton, T. F. (2015). In search of a human self-regulation system. *Annual Review of Neuroscience, 38*, 389–411.

Kober, H., Mende-Siedlecki, P., Kross, E. F., Weber, J., Mischel, W., Hart, C. L., & Ochsner, K.

N. (2010). Prefrontal–striatal pathway underlies cognitive regulation of craving. *Proceedings of the National Academy of Sciences*, *107*, 14811–14816.

Kotabe, H. P., & Hofmann, W. (2015). On integrating the components of self-control. *Perspectives on Psychological Science*, *10*, 618–638.

Le Foll, B., Di Ciano, P., Panlilio, L. V., Goldberg, S. R., & Ciccocioppo, R. (2013). Peroxisome proliferator-activated receptor (PPAR) agonists as promising new medications for drug addiction: preclinical evidence. *Current Drug Targets*, *14*, 768–776.

Levy, D. J., & Glimcher, P. W. (2011). Comparing apples and oranges: Using reward-specific and reward-general subjective value representation in the brain. *The Journal of Neuroscience*, *31*, 14693–14707.

Mayberg, H., Lozano, A., Voon, V., Mcneely, H., Seminowicz, D., Hamani, C., et al. (2005). Deep Brain Stimulation for Treatment-Resistant Depression. *Neuron*, *45*, 651–660.

Metcalfe, J., & Mischel, W. (1999). A hot/cool-system analysis of delay of gratification: dynamics of willpower. *Psychological Review*, *106*, 3–19.

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

Nitsche, M. A., Boggio, P. S., Fregni, F., & Pascual-Leone, A. (2009). Treatment of depression with transcranial direct current stimulation (tDCS): A Review. *Experimental Neurology*, *219*, 14–19.

Ochsner, K. N., Ray, R. D., Cooper, J. C., Robertson, E. R., Chopra, S., Gabrieli, J. D. E., & Gross, J. J. (2004). For better or for worse: Neural systems supporting the cognitive down- and up-regulation of negative emotion. *NeuroImage*, *23*, 483–499.

Ongür, D., & Price, J. L. (2000). The organization of networks within the orbital and medial prefrontal cortex of rats, monkeys and humans. *Cerebral Cortex*, *10*, 206–219.

Padoa-Schioppa, C., & Assad, J. A. (2006). Neurons in the orbitofrontal cortex encode economic value. *Nature*, *441*, 223–226.

Pearson, J. M., Watson, K. K., & Platt, M. L. (2014). Decision making: The neuroethological turn. *Neuron*, *82*, 950–965.

Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences*, *10*, 59–63.

Poldrack, R. A., Kittur, A., Kalar, D., Miller, E., Seppa, C., Gil, Y., et al. (2011). The cognitive atlas: Toward a knowledge foundation for cognitive neuroscience. *Frontiers in Neuroinformatics*, *5*, 1–11.

Rangel, A., & Hare, T. (2010). Neural computations associated with goal-directed choice. *Current Opinion in Neurobiology*, *20*, 262–270.

Rangel, A., Camerer, C., & Montague, P. R. (2008). A framework for studying the neurobiology of value-based decision making. *Nature Reviews Neuroscience*, *9*, 545–556.

Shah, A. K., Shafir, E., & Mullainathan, S. (2015). Scarcity frames value. *Psychological Science*, *26*, 402–412.

Smith, P. L., & Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. *Trends in Neurosciences*, *27*, 161–168.

Sullivan, N., Hutcherson, C., Harris, A., & Rangel, A. (2015). Dietary self-control is related to the speed with which attributes of healthfulness and tastiness are processed. *Psychological Science*, *26*, 122–134.

Tom, S. M., Fox, C. R., Trepel, C., & Poldrack, R. A. (2007). The neural basis of loss aversion in decision-making under risk. *Science*, *315*, 515–518.

Tversky, A., & Simonson, I. (1993). Context-dependent preferences. *Management Science*, *39*, 1179–1189.

Volkow, N. D., Fowler, J. S., Wang, G.-J., Telang, F., Logan, J., Jayne, M., et al. (2010). Cognitive control of drug craving inhibits brain reward regions in cocaine abusers. *NeuroImage*, *49*, 2536–2543.

Wunderlich, K., Rangel, A., & O'Doherty, J. P. (2010). Economic choices can be made using only stimulus values. *Proceedings of the National Academy of Sciences*, *107*, 15005–15010.

Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., & Wager, T. D. (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nature Methods*, *8*, 665–670.

Zaki, J., Lopez, G., & Mitchell, J. P. (2014). Activity in ventromedial prefrontal cortex co-varies with revealed social preferences: evidence for person-invariant value. *Social Cognitive and Affective Neuroscience*, *9*, 464–469.

Figure Captions

Figure 1. The valuation model of self-control. An arbitrary number of input processes such as primary/secondary rewards, social pressures, effort costs, and identity (left) contribute to the subjective value (middle) of the response options (e.g., the "self-controlled" and "impulsive" actions). There can be an arbitrary number of input sources and response options depending on the context and the actions that are perceived as available, and the input sources and options can change across time. The option with the highest cumulative subjective value is enacted (right). Self-control success occurs when one of the actions that align with the long-term goal accumulates the greatest amount of subjective value.

Figure 2. Value accumulation across time for two hypothetical choice options. Action A (solid line) accumulates subjective value rapidly then drops off as costs accumulate, whereas Action B (dashed line) accumulates value more slowly but eventually reaches a greater value. Action A would be selected if a low decision threshold were used because it reaches the threshold first; but Action B would be selected if a higher decision threshold were set. The selected action also depends on the timing of the decision: Action A could be selected if a time limit of 0.6 a.u. were imposed. The lack of smoothness of the lines indicates noise in the valuation process. See Smith & Ratcliff (2004) for more details on evidence accumulator models.

Figure 3. The vmPFC is involved in the computation of subjective value during choice. The region integrates heterogeneous value inputs from around the brain and computes a common currency value signal.

Figure 4. Comparison of opposition models and the valuation model. (A) Opposition models suggest that top-down and bottom-up processes compete directly with one another to drive behavior. The exact mechanism by which the competition is resolved is unspecified. (B) Valuation models state that top-down, bottom-up, and potentially many other processes each provide input into a unified value integration, which takes place in the vmPFC. The option with the highest cumulative value at the time of decision is enacted.

Figure 5. Overlap between identity and subjective value in the ventromedial prefrontal cortex (vmPFC) shown in purple. Identity-related neural activity is defined as regions active during self-processing and self-related thought (324 studies; red); value is defined as regions active during subjective value computation (812 studies; blue). Image generated using the NeuroSynth tool for automated meta-analysis of neuroimaging data (Yarkoni et al., 2011).