MULTIDIMENSIONAL DECISION MAKING

By

COLLEEN MILLS-FINNERTY, M.A.

A DISSERTATION

SUBMITTED TO THE DEPARTMENT OF PSYCHOLOGY

RUTGERS UNIVERSITY NEWARK

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN PSYCHOLOGY

WRITTEN UNDER THE DIRECTION OF

DR. STEPHEN JOSE HANSON

AND APPROVED BY

_____

_____

_____

_____

_____

NEWARK, NJ
OCTOBER 2015

ABSTRACT OF THE DISSERTATION

Multidimensional Decision Making

by COLLEEN MILLS-FINNERTY

Dissertation Director: Dr. Stephen Jose Hanson

One of the most significant questions about the nature of brain function is the extent to which brain networks are functionally constrained in terms of the task contexts in which they respond, versus whether their function is highly flexible across contexts (McIntosh, 2000). Thanks to decades of animal and human research, one of the most well defined networks in the brain is the corticostriatal circuit. The finding that dopamine neurons in this circuit respond to mismatches between expectations and outcomes in choice tasks according to a reinforcement learning (RL) algorithm (Shultz, 1997) helped orient the field of decision neuroscience towards studying RL type responses in the brain. Nearly all of these kinds of studies necessite choice tasks with *concrete* outcomes, in which subjects do something and get something, in order to have both expectation and outcome terms for RL modeling. However, far less attention has been paid to choice contexts that are *abstract,* that is, there are no expectations or outcomes. Additionally, most reward stimuli used to test decision making response are constrained by low dimensionality. This approach has left open several key questions about how the brain responds to abstract, multidimensional choices, and whether this response is modulated by factors such as choice context, subjective value, stimulus valence, and stimulus attributes in a manner that is the same or different from the well characterized response to these factors in concrete contexts.

In a series of studies, this dissertation addresses the following unresolved questions:

Study 1: How do brain regions associated with reward (e.g. striatum) and value (e.g. prefrontal cortex) respond when people choose amongst stimuli that are abstract and multidimensional?

Study 2: How is this response affected by changes in choice context and stimulus valence?

Study 3: How do value signals in the brain influence similarity coding of abstract reinforcers?

TABLE OF CONTENTS

Overview

At the broadest level, the three studies described herein aim to clarify whether the network of brain regions associated with decision making respond flexibly or if the network is constrained to primarily process only certain types of reinforcing stimuli. Thus, the first section of this introduction clarifies the reinforcer types in question, specifically the differences between concrete versus abstract reinforcers. The second section discusses what is known about the role of brain areas associated with decision making in the concrete context, and introduces key questions about the role of these areas in the abstract context. The next section aims to clarify what is known about brain connectivity during decision making, and introduces questions best addressed by using connectivity analysis as a complementary measure. The following sections present several test frameworks for comparing network response in the abstract case to the known response in the concrete case. Specifically, the role of choice context manipulations (framing), value manipulations (higher vs. lower), and valence (appetitive vs. aversive) are discussed. Then multidimensional scaling is introduced as a method for quantifying the evaluative dimensions of abstract reinforcer stimuli, and to determine whether perceptual choices for these stimuli involve value computations. Finally, an overall summary reiterates the goal of the three studies presented.

Concrete vs. Abstract

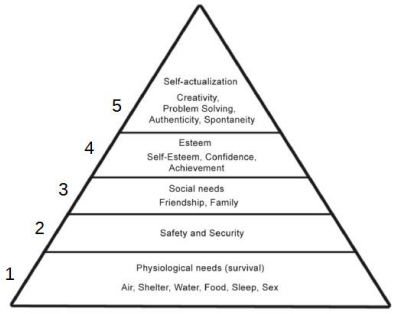
The vast majority of studies of brain response during decision making use *concrete* choice contexts, in which subjects both expect and receive an outcome based on some or all of their choices (refer to meta-analysis by Sescousse et al., 2013) . Far fewer studies examine choices for *abstract* reinforcers, which are choices that have no expectation of outcomes or actual outcomes (e.g. Sharot et al., 2009; Bray et al., 2010; Miyapuram et al., 2012; Mills-Finnerty et al., 2014). No studies prior to Mills-Finnerty et al. (2014) examine brain response to choices for stimuli that are both multidimensional and abstract. The studies herein aim to clarify whether making decisions about these complex, idiographic, hypothetical stimuli results in a similar brain response as seen in concrete choice contexts, specifically whether 1. reward circuitry such as striatum activates during choices for abstract reinforcers; 2. regions associated with value (mPFC) activate to code value of abstract reinforcers on a common scale across reinforcer types; and 3. whether abstract reinforcer decisions are modulated by stimulus valence (appetitive vs. aversive), the perceptually driven evaluative dimensions used to asses stimuli, and choice context (positive versus negative framing). Thus, the concrete versus abstract distinction is central to the aims presented here.

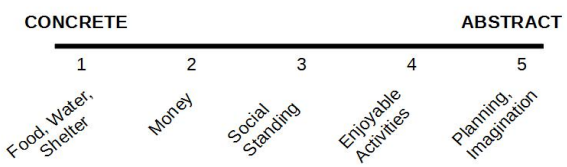
Maslow's hierarchy of needs can be used to place reward types along a spectrum ranging from highly concrete (survival needs such as food and safe shelter) to highly abstract (emotional well being). The needs at the base of the hierarchy must be satisfied before the higher level needs can be addressed (Maslow, 1943). On this continuum of reward, food represents the extreme of biological need in that it is absolutely necessary for survival and even relatively brief deprivation can lead to highly motivated behavior (see Figure 1). Money is slightly less extreme on the scale, in that lack of money does not directly threaten survival - as long as the pantry is well stocked, losing one's wallet does not have the same biological consequence as losing access to food for the same amount of time. On the next step of the continuum are needs such as social standing - in

some cultures, lack of social standing can indeed lead to threats to survival, such as in war torn countries. In developed nations, denial of this need has more pernicious long term effects, such as reduced immune function stemming from loneliness (Pressman et al., 2005). Further along the scale are the experiences of pleasurable activities such as hobbies. Denial of this need can be endured for a fairly long stretch of time; however, evidence suggests that it does have a degree of survival value. For example, in nations where leisure activities are considered a waste and work the only noble pursuit, rates of suicide are high (Targum & Kitanaka, 2012). Finally, at the most extreme end of the spectrum, most distant from immediate survival needs, are entirely abstract experiences such as daydreaming about or planning pleasurable experiences like hobbies. For people whose basic food and safety needs are met, higher order needs such as social standing or personally rewarding activities such as hobbies can be pursued. Most research has focused on choices for highly concrete rewards such as food and money. This approach has allowed the field to establish the prerequisite knowledge about decision making dynamics to make it feasible to next characterize brain response during choices for abstract reinforcers.

Several key dimensions distinguish abstract from concrete reinforcers. First, ARs meet a

Fig. 1: Continuum of Needs and Relationship to Maslow's Hierarchy





need that is situated above survival needs using the framework of Maslow's hierarchy of needs. The necessity of this is twofold: to avoid the hedonic response to stimuli like food or pain within the concrete domain, and to characterize the potentially unique processes associated with making choices about these higher-order stimuli. That is, personal enjoyment gained from a hobby is an experience likely to be a uniquely human process. Humans have capacities for memory, planning, social skills, and the ability to experience purely intrinsic reward that either is unlikely to occur or cannot be measured in other species. Since humans can verbally report how much they enjoy listening to classical music or travelling to France, this provides a unique opportunity to characterize the presumably complex series of experiences, behaviors, and judgements that lead to the manifestation of that preference.

Second, there is no experimenter determined temporal contingency between an AR choice and an eventual real-world outcome. For example, a participant making choices about types of illnesses might choose pneumonia over cancer. If that participant later contracts pneumonia, there is no contingency between performing the AR task and becoming ill. Similarly, a participant might choose between hiking and biking, and then decide to go on a hike after they leave the experimental session. The experimenter did not facilitate the hike in any way, and that time contingency is unlikely to have influenced the original choice. Importantly, in order to avoid the confound of perceived contingency, subjects should not complete the AR task multiple times. If a subject were to use the task to fulfill a stimulus-relevant goal, such as "whichever AR I pick most is the activity I'll do this weekend", then it would no longer function as an abstract reinforcer task because there would be a perceived contingency.

Abstract reinforcers are also multidimensional, in that they cannot be summarized by one or two dimensions like money (value) and food (taste/health). The many dimensions that can be used to evaluate these choices also means that choosing among abstract reinforcers relies on highly individualized processing, which enables modeling of individual differences. While money

and food hold some range of positive value for presumably all people, abstract reinforcers instead relate to individual interests in idiosyncratic ways: some people like sports while others dislike them. ARs represent needs related to self actualization and/or quality of life factors (i.e. leisure time, accomplishments within a hobby) and therefore are always in reference to an individual's likes, dislikes, and personal experiences. The degree to which these needs have been met can not be measured analogously to monetary circumstance (via bank balance) or food (via nutrition status), but studies 1b and 3 demonstrate that desired hours per week can be used as a reasonable quantification method for abstract reinforcer value.

Finally, abstract reinforcers are also hypothetical, but not merely so. A hypothetical choice for money or food would not be an AR choice because those stimuli represent a tangible stimulus tied to survival value which has a single primary evaluative dimension (value of currency). The hypothetical context is a feature that enables other AR key features such as the lack of temporal contingency, multidimensionality, and individualization of stimuli.

To summarize, the defining features of abstract reinforcers are: 1. meets a higher order need, above survival needs; 2. no temporal contingency between choices and outcomes; 3. multidimensional and individualized; 4. hypothetical. All of these criteria must be met for a stimulus or task context to meet the definition of abstract reinforcement.

Though most decisions involve evaluating multiple dimensions (such as weighing value of a gamble against the risk of losing), abstract reinforcer decisions involve a particularly high level of decision complexity, sophistication, and individual variation. Also, though these experiences do not involve tangible exchange of rewards, they nonetheless seem to be experienced as pleasurable - daydreaming of a tropical vacation provides some respite on a dreary winter day. It follows that such processes would rely on uniquely human brain architecture that allows for comprehension of concepts embedded in the idea of vacation planning, such as

"traveling," "the future" and "the self." Regions of dorsolateral and ventromedial prefrontal cortex that have been linked to memory, sense of self, and mental simulation are obvious candidate areas to serve this function (D'Argembeau et al. 2008, Roy et al., 2012; Lin et al., 2012). As mentioned previously, existing studies have used contingencies ranging from extremely concrete and linked to biological needs (receiving preferred foods, losing or winning money); to increasingly abstract such as listening to dissonant vs. melodic music or viewing pictures of car logos of inexpensive vs. luxury brands (Koelsch et al. 2006, Schaefer & Rotte, 2007); to increasingly reliant on mental simulation of potential futures, such as purchasing future vacation destinations from a fictional "travel store" (Chaudhry et al. 2009, Sharot et al. 2009). Many of these studies have found that mPFC supports the assessment of disparate rewards by coding them into a "common neural currency" (Chib et al., 2009; Smith et al., 2010; Levy & Glimcher, 2012), and that the striatum tracks the reward value of options, seemingly regardless of stimulus domain (e.g. Bartra et al., 2013). However, more research is needed to clarify whether the corticostriatal system implicated in concrete decision contexts also plays a role in choices for abstract reinforcers. Given that the primary adaptive advantage of the reward system may have been primarily to motivate behavior towards survival needs, it is possible that it would not be active during choices for abstract reinforcers..

However, another possibility follows from evidence showing that areas of the human brain that might have originally evolved for one purpose get co-opted for other, potentially more complex processes. An analogous example is the motor system, which instead of remaining limited to the execution of motor tasks, also responds when watching others perform the same task - a social skill that has obvious implications for the homo sapien first learning to start a fire. Perhaps the corticostriatal reward system has been similarly co-opted, such that making choices about needs that seem fairly distant from any survival value still rely on this system to support functions such as comparative evaluation, tracking reward value, and motivation of behavior. If

this were the case, a feedback mechanism between subcortex and cortex would likely be needed in order to integrate the complex variables that would be necessary for such a process (e.g., knowledge of one's own preferences), which is feasible given the dense anatomical connectivity between these areas.

Thus, in the three studies presented herein, abstract reinforcer stimuli are used to test whether brain network response during abstract reinforcer decisions operates under similar principles as during decisions about concrete reinforcers. The three test cases used to establish this are choice context, manipulated using a framing task; value computation, using tasks that individualize perceived high and low value of stimuli; and choice context effects on aversive abstract reinforcers, to test the effect of valence.

Brain regions involved in concrete decision making

Many recent studies focus on testing computational accounts of brain function during decision making. This approach was sparked by the findings of Shultz (1997) which indicated that dopamine neurons responded according to a temporal difference (TD) reinforcement learning algorithm during reward prediction. These types of models require information about both participant expectations and decision outcomes, and thus studies that characterize this type of response nearly always involve tasks with action-outcome contingencies (concrete reinforcement). TD and related models have been widely used to characterize the function of the brain areas involved in decision making, in particular a network including the striatum, anterior cingulate, insula, and prefrontal cortex. In this section, the current characterization of these network components in concrete decision contexts will be briefly summarized, and used as a foundation for explaining the approach that will be used to study decisions made in abstract contexts.

<u>Striatum</u>

The striatum has been the focus of the majority of cognitive neuroscience studies using TD models of decision making (Bartra et al., 2013). Most studies have found that striatal activation goes up for outcomes that are better than expected (positive prediction error) and down for outcomes worse than expected (negative prediction error; e.g. Knutson et al. 2003; Delgado et al., 2003; Tom et. al, 2007; Pessiglione et al., 2006). However other studies have found increases in activation for both unexpectedly positive and negative outcomes suggesting a broader salience or novelty coding mechanism (e.g. Cooper & Knutson, 2008). Striatum and VMPFC are also involved in coding prediction errors where the comparison is between one's own preferences, and the preferences of another person, suggesting that the PE framework is fairly flexible (Garvert et al., 2014). All such processes rely on generating an error signal based on concrete values for both expectations and outcomes of choices. It is unclear how this prediction error account applies to decisions made when there are no tangible expectations because subjects are not, for example, expecting to be rewarded with the outcome of one choice during a randomly chosen trial, and there is no outcome phase in which a choice results in a consequence. One hypothesis would be that the striatum would not be active during these decisions since it normally codes for these factors. However, if the striatum is active, this raises the possibility that the striatum tracks a broader range of stimuli than has been characterized by previous research. This would suggest that the definition of reward should be expanded to include not only anticipated actual experiences, but also the abstract desire for the experience of pleasurable activities.

Anterior cingulate

The anterior cingulate (ACC) is implicated in a wide range of functions ranging from fear conditioning and extinction, emotional conflict and regulation, and cognitive control of attention and memory (for review and meta-analyses see Etkin et al., 2011, Shackman et al. 2011, Gasquoine 2011, Niendam et al. 2013). While there is general agreement that this region is a densely connected hub involved in many different processes, there is disagreement about the

specific factors that influence its response. One theory, the predicted response outcome model, suggests that it's role can generally be described as flexibly predicting outcomes of actions regardless of valence (Alexander & Brown, 2011). Another theory, the expected value of control theory, incorporates both outcome prediction and a cost function sensitive to the effort required for a task, also independent of valence (Shenhav et al., 2013).

The key difference between these theories of ACC function is functional specificity - whether ACC supports general emotion regulation and cognitive control, or whether this response is specific to contexts involving action sequences in pursuit of a goal. If ACC is using the outcomes of past actions to inform future actions, it should only be active and functionally connected to the decision making network in tasks where there are action-outcome contingencies. If it is more flexibly involved in maintaining attention and regulating emotion, it should be active during abstract decisions, and be modulated by factors such as decision context being positive or negative and value or salience of choices.

Insula

Computational approaches have also been used to characterize function of the insula. Most generally, insula activation has been shown to be related to interoceptive processing and awareness of internal states, both physical and emotional. The insula is active in response to stimuli that provoke disgust, while participants reject unfair offers in the ultimatum game, observe others experiencing pain, and compute prediction error for aversive food stimuli (Wicker et al. 2003, Sanfry et al. 2003, Kim et al. 2006, Singer et al. 2009). Although less research focuses on its role in processing positive states, insula activation is also observed when both experiencing and viewing others tasting appealing drinks, during pleasant touch, and during orgasm (O'Doherty et al. 2001, Komisurak et al. 2004, Ortigue et al. 2007, Singer et al. 2009).
Some theories suggest that the insula integrates conscious appraisal with bodily feelings to facilitate general awareness of emotional states (as opposed to unconscious affective reactions,

such as to masked negative stimuli, which seem to be primarily processed in the amygdala; Morris, 2002). Other accounts suggest that the insula helps integrate information crucial to decisions, such as processing bodily signals/homeostatic states and translated them into urges to act, particularly in the case of drug craving but also in scenarios such as risky gambling (Xue et al. 2010, Nacqvi et al. 2007, Nacqvi & Bechara 2009). Another model suggests that the insula integrates information about uncertainty, physical states, personal preferences, and contextual appraisal to inform decision making (Singer et al., 2009).

These theories differ regarding whether they assert that insula response is specific to concrete contexts or if it performs a more general role in integration of factors like affect and physical states. Many task contexts designed to evoke insula response have contingencies by default, such as evaluating food rewards or monetary offers in game situations. If the insula codes generally for emotional awareness rather than computing an outcome based variable such as risk, then it should be active during decisions about abstract reinforcers, and may parametrically track factors such as stimulus intensity.

Prefrontal Cortex

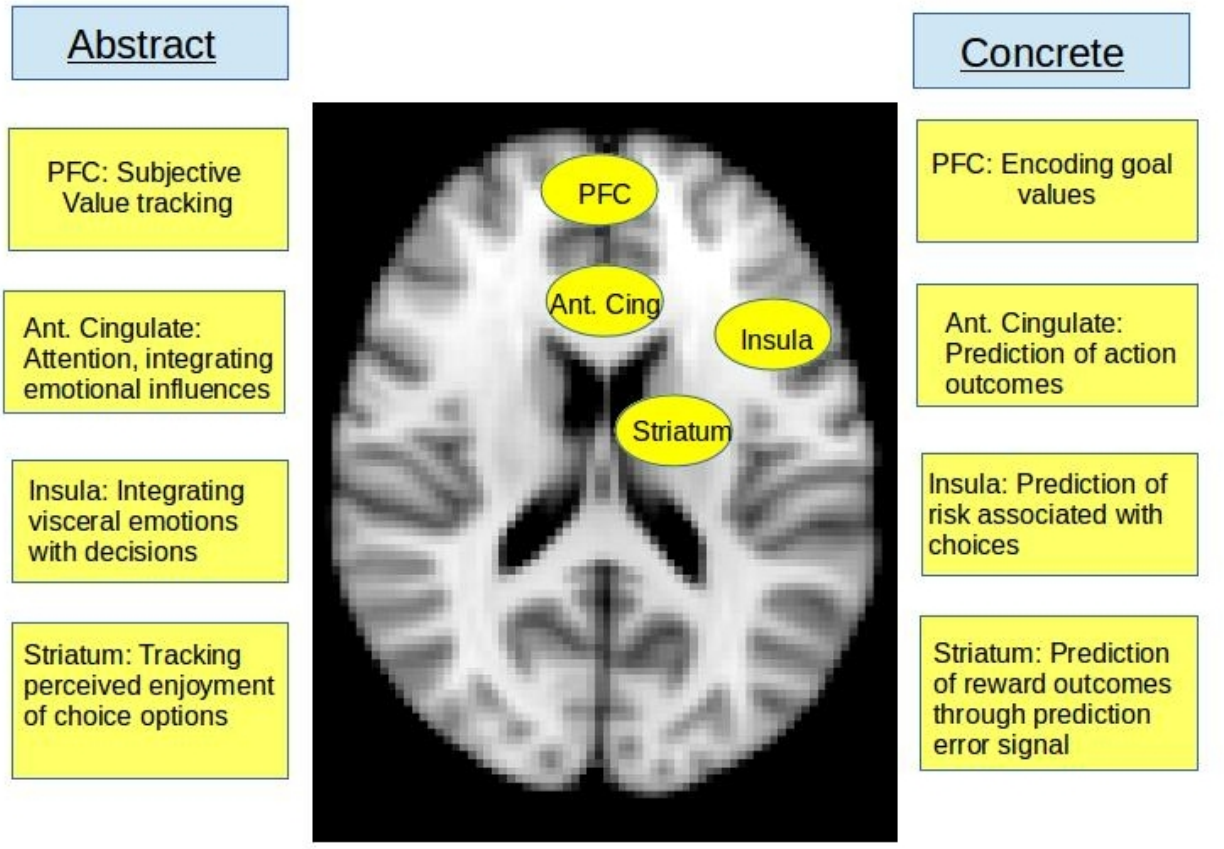
The ventromedial prefrontal cortex (mPFC) is thought to play a critical role in the consolidation of information underlying subjective value judgements across domains in order to facilitate goal directed action (e.g. Chib et al. 2009, Grabenhorst & Rolls 2011). This is likely not the exclusive function of mPFC, as it has also been implicated in abilities as diverse as memory, default mode processing, and emotion (Roy et al. 2012). However, there is convincing converging evidence that computing value is a unique and primary function of the region. mPFC has been implicated in choices involving monetary options (including probabilistic and delayed rewards), food, brands, music, colors, vacation destinations, consumer goods, preferred faces, and social rewards (Johnson et. al, 2005, Knutson et al. 2007, Kim et al. 2007, Lebreton et al. 2009, Peters & Buchel 2009, Hare et al. 2009, Sharot et al. 2011, Qin et al. 2011, Santos et al. 2011,

Lim et al. 2013). mPFC activation magnitude changes dynamically in response to value, increasing along with positive expected outcomes and decreasing in magnitude of activation to choice options that have been devalued experimentally (Kim et al. 2006, Qin et al., 2011). It has also been proposed that mPFC plays a unique role in transforming subjective value judgements into choice behavior, though whether this is accomplished through a computational approach that coordinates with the motor system or a feedback based approach involving subcortical areas remains a topic of debate (see Reimann & Bechara 2010; Grabenhorst & Rolls, 2011).

Though many studies have posited that mPFC codes dissimilar choice options into a "common neural currency" (e.g. Izuma et al. 2008, Peters & Buchel 2009, Levy & Glimcher 2012), thus explaining its involvement in a wide range of choice behavior, several questions remain. Some studies have tested mPFC response across different task contexts such as risky versus ambiguous rewards (Levy et al. 2010) or motor vs. visually cued rewards (Glascher et al. 2009), while others have varied the choice parameters to include a range of concrete rewards like money, food, or trinkets (Chib et al. 2009). However, no studies prior to Mills-Finnerty et al. (2014) varied choice parameters to include a range of abstract preference categories in a within subjects design. Many studies also only present preferred stimuli pre selected by participants, such as vacations or paintings they ranked as highly preferred, or use stimuli that only include pairs of equivalently ranked options as seen in many cognitive dissonance studies (Chaudhry et al. 2009, Jarcho et al. 2009, Qin et al. 2011). Finally, studies generally offer a constrained set of rewards without clarifying whether those rewards are perceived as equally valuable to all participants - for example, offering two types of food (sweet and salty) as high value food reward while not controlling for whether all participants actually find those options high value (Levy & Glimcher, 2011). These issues, along with methodological confounds of many value studies, are discussed in more detail in Section 6, Valuation processes.

In sum, while the role of brain regions associated with decision making is reasonably well established for concrete reinforcer choices, many open questions remain regarding the role of these regions during abstract reinforcer choices, ranging from whether they respond at all to if they do, what this response looks like. We hypothesize that abstract reinforcers engage the same brain areas as concrete reinforcers, but the role these brain areas play will be different due to the different task demands for making decisions about abstract reinforcers. Specifically, as there are no explicit goals when choosing between abstract reinforcers, we expect that activation of the prefrontal cortex will track the perceived or subjective value of the options presented rather than tracking goals as in concrete decision making. The anterior cingulate, rather than predicting outcomes of choices, since there are no outcomes, will instead direct attention and integrate affective feedback to guide choice behavior. The insula, instead of predicting risk, will integrate affective components such as how salient or arousing choice options are. Finally, the striatum will not track reward outcomes via a prediction error algorithm requiring terms for expectations and outcomes, but will instead track the perceived reward value or attractiveness of the choice options in order to guide decisions. Figure 2 outlines the role of these regions in concrete contexts, and hypotheses for the equivalent function in the abstract context.

Fig.2: Role of brain regions in abstract vs. concrete decisions



Connectivity

The majority of decision making studies use a General Linear Model approach to test predictions about data, in which changes in magnitude of activation are the primary outcome measure. However, there is growing appreciation of magnitude-independent signal in the brain. That is, there may be areas of the brain active at magnitudes below standard statistical thresholds whose activity is nonetheless tightly coupled with other regions in a task relevant fashion. Furthermore, high magnitude activation that survives statistical thresholds may not represent direct communication dynamics but rather independent foci of activation. However, recent methodological advances enable the study of brain connectivity (Smith et al., 2011). A range of approaches including functional connectivity and Bayesian modeling methods aim to quantify the interactive, communicative dynamics amongst different regions that comprise brain networks.

This approach is particularly valuable for clarifying the role of regions such as the insula or anterior cingulate, which are active in a wide range of tasks in a seemingly context dependent manner. That is, the same magnitude of activation in insula in task Condition A versus Condition B may signify different processes, incoming signal(s), and outcoming signal(s). Furthermore, these differences cannot be tested using GLM as equivalent magnitude of the activation in both conditions means that contrasts (e.g. using a t-test) will produce what appears to be a null result. Thus, understanding connectivity in addition to magnitude effects is of great relevance to understanding brain processes.

Many studies address these issues by using correlation among brain areas to define networks; however, analyzing co-active voxels can lead to false positives (particularly given the sheer number of voxels included in whole brain analyses, often 10,000+), and does not provide information about the orientation of connections. Other connectivity methods that do specify connection orientation also have limitations such as having to prescribe a model in advance, as with structural equation and dynamic causal models, or the method not being developed with BOLD data in mind and being potentially susceptible to vascular influences, as with Granger Causality (Webb et al., 2013). Connectivity analysis using Bayes networks provides information about network response including: the direction of temporal influence among nodes (ROIs) in the network, the strength of the connections (edges) between nodes, and the number of edges coming in or out of a node. Since the method is search based, a model does not need to be prescribed in advance, and since the search is constrained to pre-chosen ROIs and penalized for overfitting, this helps reduce the risk of false positives substantially. In studies 1 and 2, we conducted our connectivity analysis using a Bayes net method called IMaGES (Independent Multi-Sample Greedy Equivalence Search) algorithm (Ramsey et al. 2010, 2011). With simulated data, IMaGES performs at 98% recall and 90% precision (Ramsey et al., 2010, 2011). IMaGES results also show concordance with anatomical connectivity patterns measured using DTI (Sun et al., 2012).

IMaGES has been used for network analysis in tasks ranging from semantic (Boukrina & Graves, 2013), to social brain network response in patients with autism (Hanson et al., 2013), to working memory (Manelis & Reader, 2014). More detail on the implementation of this method can be found in the methods sections of Studies 1a and 2.

A growing number of recent studies have used connectivity analysis to examine patterns of interaction between brain regions during decision making. One study found that during risky gambles, the anterior cingulate was functionally connected to regions such as the striatum, OFC, and amygdala that are frequently implicated in decision making (Cohen et al., 2005). Another study compared the effects of a diet goal on decisions for food, and found both DLPFC downregulation of inferior frontal gyrus, and positive connectivity between IFG and mPFC when dieters rejected unhealthy foods (Hare et al., 2009). These results suggest that different regions of PFC may be functionally specialized for different parts of the decision process. Specifically, DLPFC is hypothesized to be involved in executive control and goal orientation whereas mPFC is thought to code a value signal that incorporates those factors. Similar results come from a study of probabilistic gambles, in which both frontopolar cortex activation and connectivity patterns with parietal and premotor regions preceded decisions to switch choice strategy, suggesting a role for that frontopolar cortex in evaluating alternative courses of action (Boorman, 2009).

Other studies have used connectivity to delineate the areas providing input to mPFC in order to facilitate value coding. One study found that during a task in which participants assigned value to choice options based on visual and semantic features (color and logo on tshirts), functional connectivity occurred between fusiform gyrus and mPFC during the visual condition, and between posterior superior temporal gyrus and mPFC during the semantic condition (Lim et. al, 2013). These results support the idea that specialized areas process relevant attributes of stimuli which are then integrated in mPFC into an overall value signal. Similarly, in a study of moral judgments, amygdala connectivity with mPFC changed as a function of the aversiveness of

choices (Shenhav & Green, 2014). In another study examining value coding of attractive faces, temporoparietal junction and middle temporal gyrus connectivity with mPFC was related to subjective value judgements, suggesting that the social relevance of stimuli requires computation in areas associated with social cognition before value can be computed in mPFC (Smith et al., 2014). A related study had subjects make choices to either keep monetary wins for themselves, or split those wins with another person. Participants were also given information about the social closeness of the hypothetical other person, to test how social distance influences such choices. The authors found that TPJ-mPFC connectivity was related to the "social distance dependent balance between generous and selfish motives;" that is, overcoming the impulse to be selfish in order to facilitate generosity to a close other is related to the strength of this connection (Strombach et al., 2015).

Connectivity can also be used to substantiate mechanistic explanations of brain function related to anatomical connectivity and neurotransmitter function. For example, one study used connectivity analysis to clarify the role of different types of salience processing and its effect on reaction time during value based decision making (Kahnt & Tobler, 2013). They found that connectivity between the temporoparietal junction and locus ceruleus was related to salience coding. These results suggest a mechanism for how noradrenergic projections in the locus ceruleus help govern attentional mechanisms in temporoparietal junction. Another study examined the response of dopaminergic circuitry while subjects listened to classical music, and found connectivity between VTA, NA and hypothalamus, suggesting that the experience of listening to music involves dopaminergic reward circuitry in the brain (Menon & Levitin, 2005).

Each of these studies suggests that the areas frequently active during decision making also interact with each other by playing different roles depending on decision making context. While in aggregate (e.g. meta-analyses), GLM based results have suggested this to be the case, connectivity analysis adds a further level of clarity by quantifying *how* these regions interact,

such as by measuring the directionality of connections. In several of the studies herein, we use a Bayesian heuristic search method (Independent Multi-sample Greedy Equivalence Search; Ramsey et al., 2010, 2011) in order to quantify these dynamics, particularly as they pertain to the insula and anterior cingulate. Specifically, in Study 1 we test the hypothesis that during negatively framed abstract reinforcer decisions, longer reaction time is related to increased affective processing particularly in terms of the strength of connectivity between anterior cingulate, insula, and striatum. In Study 2 we test whether the amygdala contributes to value coding of aversive stimuli through stronger or an increased amount of connectivity with areas such as striatum and prefrontal cortex.

Framing Effects on Decision Making

Behavior and brain activation during decision making are frequently modulated by choice context. Context can refer to any number of manipulations such as varying the number of choices presented at once, time delays associated with choices, whether choices result in losses, gains, or both, etc. Often these manipulations test how context leads to behavioral bias, such as a large choice set reducing the likelihood of a purchase decision compared to a smaller choice set (Iyengar & Lepper, 2000). One way to test whether concrete and abstract reinforcer contexts rely on the same or different corticostriatal mechanisms is to investigate whether the decision making biases seen in concrete contexts also occur when subjects choose between abstract reinforcers. Given the increasing number of studies of contextual effects, a full discussion of the topic is beyond the scope of this section. Instead, several illustrative examples of cognitive neuroscience studies of contextual effects on decision making are discussed, and then an approach for testing abstract reinforcer responses using framing effects is discussed in more detail.

One well characterized decision making bias is delay or temporal discounting, the tendency to prefer a smaller reward now over a larger reward later. Studies suggest that individual differences in bias towards smaller immediate rewards is related to greater limbic activation,

particularly in striatum (McClure et al. 2004; Harari et al., 2006; Luo et al., 2009; Peters & Buchel, 2009; Ballard & Knutson, 2009). Delay discounting is also related to impulsivity as a personality trait (Odum, 2011), particularly in addiction populations (refer to meta-analysis by MacKillop et al., 2012). Choices for both real and hypothetical stimuli are vulnerable to this bias (Odum et al., 2006; Bickel et al., 2009). There is also increasing evidence suggesting that although delay discounting behavior tends to be stable long term, interventions involving working memory or episodic future thinking can change discounting behavior (Bickell et al., 2011; Radu et al., 2011) as well as related brain activation (Peters & Buchel, 2010).

Social context can also bias decision making and have effects on brain activation. For example, when participants believed that neutral consumer items were preferred by members of an outgroup (the opposing political party), they devalued those items during choice, which was related to decreased VMPFC activation (Kim & Johnson, 2014). Items associated with the in-group were more likely to be chosen and were related to increases in VMPFC activation. Social context effects can also interact with other biases - for example, in adolescents performing a delay discounting task, smaller immediate rewards are preferred more in the presence of a same age peer, regardless of whether or not the peer is familiar or a stranger (Weigard et al., 2014). The discounting framework can also be applied to social contexts more broadly, in which social distance governs the degree of empathy or selfishness for others (Strombach et al., 2014).

Importantly, biased behavioral results can occur even when context manipulations appear irrelevant to the task at hand. For example, a novel but non-predictive image cue led to increased hippocampal and striatal activation for gambles with moderate (but not high) probability, suggesting an interaction between cue novelty and reward anticipation during ambiguous choice (Guitart-Masip et al., 2010). Additionally, rating items after making choices about them can also bias decision making. When subjects are forced to choose between two equally preferred items, preference ratings for the chosen item tend to increase whereas the rejected item is devalued,

which in turn affects future choices. Several studies have found that these revaluation processes are related to mPFC activity (Jarcho et al., 2010; Qin et al., 2011). Other recent studies have found that irrelevant stimuli that precede a choice prompt, such as pleasant music, can then lead to biases in subsequent decisions for stimuli with variable subjective value such as paintings (Abitbol et al., 2015). This may also work in reverse; that is, when subjects make value judgements, their subjective value assignment is related to behavior and brain activation during a subsequent perceptual decision making task during which value is irrelevant (Grueschow et al., 2015).

One contextual bias that is particularly well established is the framing effect. Studies of this effect have shown that even when the actual outcomes are equivalent, whether a choice is framed as a gain or a loss will alter the behavioral response or decision made (e.g. Kahneman & Tversky 1986; refer to Kuhberger 1998 for meta-analysis). In the seminal Kahneman and Tversky study of the framing effect, they found that people make different decisions about a hypothetical vaccine when "200 people will be saved" versus "a ⅓ probability that no one will die and a ⅔ probability that 600 people will die" even though the outcome is the same in both scenarios (Kahneman & Tversky 1981). In a more recent neuroimaging study, a $0 lottery winning was perceived as aversive if the other options are winning $2 and $5, but appetitive if the other options are losing $2 and $5 (Brieter et al., 2001). The mechanisms of this bias are thought to be as follows: loss frames tend to encourage riskier decisions than gain frames, due to loss aversion, whereby offsetting a loss requires a gain twice as large. Under the threat of loss, riskier decisions become more appealing if they offer the chance at avoiding a loss altogether. This bias exists in a range of domains including gambling tasks, choices about health outcomes, consumer product decisions, and solving logic problems (Kahneman & Tversky 1986, Biswas & Grau 2008, Rothman & Salovy 1997, Kuhberger 1998, DeMartino et al., 2006).

In terms of brain response, areas implicated include the amygdala (DeMartino et al., 2006), striatum and ventromedial prefrontal cortex (Tom et al., 2007; Foo et al., 2014), and dorsolateral prefrontal cortex (Gonzalez et al., 2005; Foo et al., 2014). In one study using monetary gambles, increased activation in orbital and medial prefrontal cortex was correlated with decreased susceptibility to framing, meaning less bias towards risky decisions during loss frames (DeMartino et al., 2006). Behavioral and neuroimaging studies generally tend to demonstrate longer reaction times during negative contexts compared to positive (Kim et al. 2006, Fitzgerald et al., 2009, C. Alos-Ferrer et al., 2012). Additionally, patients with autism showed reduced responsiveness to framing manipulations, which was related to decreased emotional response to negative frames and a lower degree of risk aversion compared to controls (De Martino et al., 2008). Participants making judgements about self relevant descriptors such as cleverness or honesty were more likely to endorse positively framed statements (i.e. "I am honest at least 75% of the time") than negative ("I am not honest up to 25% of the time"). Positively framed judgements were related to greater mPFC activation, whereas negative judgements activated regions such as the insula that process aversiveness (Murch & Krawczyk, 2014). Subjects playing a "double or quits" gambling task were more likely to choose riskier second gambles after a prior loss, than after a gain. VMPFC responded to both risky gambles in a loss context and safe gambles in the gain context, suggesting a contextually governed response (Losecaat Vermeer et al., 2014). Finally, a study of go/no-go behavior presented cues representing an aversive or appetitive juice stimulus prior to trials, and found that appetitive cues biased the motor system towards movement, whereas aversive cues bias it against movement. This bias was related to increases in go vs. no-go errors, and the direct effect of appetitive versus aversive cues on motor system excitability was confirmed using TMS (Chiu et al., 2014). These studies all suggest that framing can affect a range of concrete reinforcer decisions including money, self relevant descriptors, and avoidance learning.

Thus, since brain response to framing manipulations is well characterized in the concrete domain, the studies herein test whether framing affects abstract reinforcer decisions. Specifically, abstract reinforcer choices are framed positively/in terms of gain, versus negatively/in terms of loss. Within the appetitive domain, the positive or gains frame used is "which do you like more" versus "which do you like less" for the negative or loss frame. Within the aversive domain, the positive frame is "which would you rather avoid" since in this case avoiding a negative outcome is positive; this is referred to as the avoidance frame. Choosing "which would you rather have" when both choices are aversive results in a loss focused or negative frame, referred to as the approach frame. For both appetitive and aversive abstract reinforcer choices, we predict longer reaction for negative frames. Negative framing of choices is also predicted to result in greater activation and connectivity amongst limbic areas such as striatum, insula, and amygdala. More specific hypotheses related to framing can be found in the introductions for Study 1a and Study 2.

Value

A large body of recent research has supported a role for the medial prefrontal cortex (mPFC) in tracking subjective or perceived value (Grabenhorst & Rolls, 2011; Roy et al., 2012; Levy & Glimcher, 2012; Bartra et al., 2013; Clithero & Rangel, 2014). mPFC has been implicated in assessing value of a diverse array of choice options, including both real and hypothetical monetary rewards, food, brands, music, colors, vacation destinations, consumer goods, fractals, preferred faces, and emotional and social rewards (Johnson et. al, 2005, Knutson et al., 2003, Kim et al., 2007, Lebreton et al., 2009, Peters & Buchel, 2009, Hare et al. 2009, Bray et al., 2010, Sharot et al., 2009, Qin et al. 2011, Santos et al. 2011,Miyapuram et. al 2012, Litt et al., 2012, Lim et al. 2013, McNamee et al., 2013, Winecroft et. al, 2013) Foo et. al, 2014, Gross & Woelbert, 2014, Smith et al., 2014). Many of these studies have found that mPFC supports the assessment of disparate rewards by coding them into a "common neural currency" (Chib et al., 2009; Smith et al., 2010; Levy & Glimcher, 2012). In addition to a role in value coding, mPFC

has been implicated in abilities as diverse as memory (e.g. Barron et al., 2013), self relevance (see meta-analysis by Northoff et al., 2006), default mode (e.g. Gusnard et al, 2001), social context (e.g. Kim & Johnson, 2014), mental simulation (e.g. Benoit et al., 2014), and emotion (Roy et al. 2012; Lindquist et al., 2014). Results from non-human primate studies suggest that some mPFC neurons are specialized for coding the value of rewards (e.g. Tremblay & Schultz, 1999 ;Wallis & Miller, 2003), but neurons in similar regions can also shift their response based on context, such as social ranking of conspecifics (Azzi et al., 2012). A recent review of rat literature also suggested that "medial and orbital prefrontal regions frequently respond very differently to the same experience in the same brain and the rules that govern prefrontal plasticity appear to differ for those of other cortical regions" (Kolb & Gibb, 2015). Recent meta-analytic results in humans suggest that different mPFC connectivity sub-networks representing a posterior to anterior gradient of response to rewards ranging from concrete to abstract (Clithero & Rangel, 2014) and recent study results suggest this gradient may be bounded by individual anatomy (Li et al., 2015). However, much is yet to be learned about the organizational rules that govern mPFC function in both animal and human models. Specifically, one of the most pressing unresolved questions about mPFC function is whether specialized subregions exist, or if the same cell populations in this region are capable of processing a wide range of stimuli as suggested by results from animal studies.

Further, several important methodological concerns highlight limitations of the inferences that can be made from human neuroimaging studies of value (O'Doherty, 2014). Specifically, stimulus value is often inherently confounded with stimulus salience, sensory encoding of stimuli, self-referential processing, goal motivation, familiarity, or a combination thereof. This makes it difficult to determine whether voxel activation in mPFC is truly value related versus driven by these other factors. The following section discussing these confounds and how they can be partially or completely controlled for by using abstract reinforcer stimuli.

Goal motivation related to action-outcome contingencies ("you do something and you get something") can potentially drive brain response at or around the same time value computations are being made (e.g. O'Doherty, 2011; Glascher et al., 2012). Even simple decisions have several component processes - the subject must orient to stimuli, evaluate the options, make a selection often involving a motor sequence like button press, potentially wait for an outcome, and then receive an outcome. Recent meta-analyses suggest that the latter four phases may activate different regions of mPFC (Clithero & Rangel, 2014). Even in studies that do examine hypothetical choices, there are often still inherent goal motivations tied to stimuli such as maximizing payout, or mentally simulating the experience of receiving a monetary reward (Miyapuram et al., 2012). In Study 1b we examine preference choices where there are no implied, simulated, or actual outcomes, an approach that may help ascertain the degree of functional localization of mPFC response to value in non-contingency choice contexts.

Salience and sensory attributes of stimuli can also vary along with value (O'Doherty, 2014). In a typical task, options that are higher in value are often also inherently more arousing or salient (e.g. highly appetizing foods, gambles with a higher payoff). Outcome identity and outcome value might also be collinear, such that there are physical properties of the stimulus (taste, smell, color) that could drive a stimulus-encoding signal in mPFC which could appear identical to value coding. Here, by individualizing the stimuli presented to subjects, we avoid the confound of sensory features of stimuli. Since stimuli also vary in dimensions related to salience (such as intensity or novelty), this factor is at least partially controlled for as well.

Studies have also found that cortical midline structures, particularly mPFC, are involved in processing self referential stimuli across social, emotional, and other domains (Northoff et al., 2006). Choices made in a self relevant context elicit more prefrontal activation compared to non-self relevance choices (Johnson et al., 2005). Other studies have found that moral value choices related to self concept activate different parts of PFC than value judgements for concrete actions

(Brosch et al., 2012) and that mPFC tracks response to self relevant stimuli according to social context (Kim & Johnson, 2014). While self referential processing may be a component process or input used in coding value of stimuli like favorite hobbies, we feel it is important to disentangle these factors. For example, it may be that the self referential factors in a decision are coded in one part of mPFC, but the overall value signal is coded in another. In Study 1b, we ask participants to rate the relationship between preferred stimuli and their self concept to use this variable as a control measure.

Finally, controlling for familiarity is a significant issue when studying preference decisions. Decades of studies of the mere exposure effect (Zajonc, 1968) have demonstrated that simply exposing participants to stimuli increases liking ratings. More recent studies suggest that for stimuli such as music, familiarity drives a significant amount of brain response to preferred songs (Pereira et al., 2011). Many studies sidestep this issue by presenting novel stimuli such as fractals; however, doing so involves sacrificing some of the ecological validity of stimuli, as the things we like most in everyday life are typically things we're familiar with. Since participants in Studies 1b make choices about stimuli they have real life involvement with, we are able to quantify this familiarity and use it as a control variable.

In sum, mPFC presents one of the larger challenges in mapping both local and network function in the brain, in that many different processes likely co-occur in the same voxel sets and possible even the same neurons. Furthermore, separating out these processes presents a series of methodological challenges. However, since abstract reinforcer stimuli vary along many dimensions including self reference, familiarity, inherent salience, and sensory attributes, they provide a unique framework for possibly overcoming some of these challenges in order to clarify value specific responding in mPFC.

The role of valence in decision making

While the role of brain networks in decisions for appetitive concrete reinforcers is fairly

well established, it is less clear how the corticostriatal circuit responds to aversive reinforcers. Several possibilities exist: the same regions may code both appetitive and aversive stimuli by proportionally increasing or decreasing their activation relative to the positive or negative value of stimuli; an overlapping set of regions may code common aspects of both appetitive and aversive choice scenarios (such as goal motivations) while the actual aversive elements are processed in potentially more specialized limbic regions such as the insula and amygdala; and a final and less likely possibility is that the brain networks for aversive and appetitive decisions consist of entirely separate sets of brain regions. Recent meta-analysis suggests that the second hypothesis is most strongly supported by the current literature (Hayes et al., 2014; Lindquist et. al, 2014, Liu et al., 2011). However, few studies have examined whether brain response when experiencing an aversive stimulus is different from choosing amongst hypothetical aversive stimuli (i.e. Sharot et al., 2010; Feldmen-Hall et al., 2012; Kang & Camerer, 2013). None have examined response to heterogeneous types of abstract reinforcers, or characterized differences between approaching and avoiding such stimuli, a gap addressed by Study 2.

Another key question is whether brain response to aversive stimuli is driven primarily by salience or valence. The salience account predicts increases in activation for all highly arousing stimuli, regardless of valence, whereas most valence coding accounts hypothesize that appetitive stimuli should lead to parametric increases in activation in response to increasing reward value, and aversive stimuli should result in parametric *decreases* (Lindquist et al., 2014). Results supporting the former have been observed across many studies in areas such as the striatum and VMPFC (refer to meta-analysis by Sescousse et al., 2013), but results in favor of the latter remain mixed. For example, one recent study found that signal in PFC for both appetitive and aversive stimuli could be identified using multivariate pattern analysis (MVPA; Kahnt et al., 2014). However, since MVPA measures overall patterns of activation that may be a summation over increase and decreases, these results do not clarify whether the magnitude change for aversive

stimuli is in the positive or negative direction. Studying deactivations poses a number of methodological issues. For example, decreases in activation can only be measured relative to a baseline, usually activation during a control condition or rest, which limits interpretation across studies due to protocol differences. Further, comparing deactivations directly (e.g. whether an area reduces activation more in one condition versus another) poses a challenge in terms of basic modelling which is not addressed by most commonly used software packages (such as by allowing comparison of absolute values). In fact due to a lack of deactivation based results as well as these interpretive challenges, no studies of deactivation were included in a recent meta analysis of aversive brain response (Lindquist et al., 2014). However, one approach that can uniquely address these issues is connectivity analysis, because it is agnostic to magnitude; that is, because it measures interactions amongst regions using time series data, it doesn't matter if those interactions involved significant increases or decreases in activation relative to baseline. Studying whether the number, strength, or direction of connections changes in "deactive" regions can therefore uniquely characterize these brain dynamics. Thus, in Study 2 both deactivations and activations in response to aversive stimuli are identified using GLM, and connectivity analysis is then used to compare deactivation dynamics across conditions.

According to several recent meta-analyses, areas that may be specialized for processing the value of aversive stimuli include posterior cingulate, amygdala, parahippocampus, and inferior frontal gyrus (Liu et al., 2011; Hayes et al., 2014; Lindquist et. al, 2014). Areas selective for appetitive stimuli include anterior cingulate and superior temporal gyrus. Areas that may play a role in both include thalamus, amygdala, hippocampus, insula, ventral striatum, and certain regions of VMPFC and DLPFC. Studies included in the Hayes et al. (2014) meta-analysis involved passive exposure to aversive stimuli, and therefore cannot account for learning or habituation effects. Studies of painful stimuli were excluded to avoid the "unique subjective experience associated with these stimuli" (Hayes et al., 2014). The Lindquist et al. meta-analysis

(2014) focused on positive and negative emotional reactions to stimuli such as emotional facial expressions. Since valence and salience are collinear for stimuli used in the majority of such studies, it is not possible to determine whether the areas they identified as generally coding valence are truly specific to valence. They acknowledge this limitation, as well as challenges associated with the lack of studies that use both aversive and appetitive stimuli within the same paradigm (Lindquist et. al, 2014). Study 2 addresses unresolved questions about how value of abstract reinforcers is coded by utilizing aversive stimuli, and also provides a framework for examining both activation and deactivation related to valence using a combination of General Linear Model and connectivity analysis.

More generally, human neuroimaging studies have established brain response to a range of aversive stimuli, including unappealing foods or beverages (Harris et al., 2011; Kang & Camerer, 2013; Metereau et al., 2014), negative feedback (Bhanji & Delgado, 2014), electrical shocks (Collins et al., 2014; Lawson et al., 2014; Winston et al., 2014) monetary losses (e.g. Delgado et al., 2003; Kahnt et al., 2014), tactile stimulation (e.g. uncomfortable heat, pressure, or textures; Roy et al., 2014; Lamm et al., 2014), odors (Gottfried et al., 2002), unattractive faces (Martín-Loeches et al, 2014) and illnesses (Sharot et al., 2010). Most studies examine response to one stimulus type. Fewer studies look at multiple types of aversive stimuli in the same subjects (e.g. Lamm et al., 2015; Metereau et al., 2014), and none to our knowledge customize stimuli across multiple categories based on participant perceptions of aversiveness as is done in Study 2.

Several recent studies examine how aversive stimulus decisions interact with other behaviors. One study examined how uninstructed emotion regulation affects response to aversive IAPS pictures (Silvers et al., 2014). They found that increased DLPFC and VMPFC activation in concert with decreased amygdala activation was related to lower levels of negative affect during viewing of negative images. These regions are frequently implicated when participants are trained or directed to regulate emotion, and the results suggest that individuals vary in their propensity

and "natural" ability to regulate negative emotional responses. Another study used juice, money, and pictures as both appetitive and aversive stimuli, and found increased VMPFC response to both highly appetitive and highly aversive options, suggesting an "unsigned" value signal or salience driven response (Metereau & Dreher, 2014). Another study examined pleasant and unpleasant visuo-tactile combinations, such a picture of a toad paired with the application of a slimy substance to a participant's hand while they were in the scanner. They found that pleasant combinations, such as the image and sensation of silk, was related to increased OFC activation, whereas negative combinations were processed in an insular-frontal circuit (Lamm et al., 2015). Another study examined threat avoidance behavior by having subjects engage in a task where if they did not move their cursor within an allotted timeframe, they would receive a mild electrical shock (Collins et al., 2014). They found that stronger coupling between amygdala-VMPFC and caudate-VMPFC predicted better avoidance learning by the end of their task. While these studies address interesting interactions between valence and other factors such as avoidance learning and emotion, no studies thus far have addressed avoidance versus approach behavior with aversive abstract reinforcers. It is unclear if making choices amongst these stimuli relies on the same brain mechanisms as for concrete reinforcers, particularly whether avoiding a hypothetical aversive scenario is processed in the brain similar to positive reward, a limitation addressed by Study 2.

A small number of studies have examined choices for hypothetical aversive stimuli. One study found that monetary bids to avoid aversive food stimuli were higher for real choices compared to hypothetical, which was in turn related to greater amygdala and insula activation to risky bids in the real vs. hypothetical case (Kang & Camerer, 2013). In another study where subjects could pay from their own pool of experimental money to reduce the severity of shocks they thought were administered to a confederate, participants made higher monetary offers when they imagined the confederate being shocked (hypothetical condition) and lower offers when they saw what they thought was live video of the shocks ("real" condition). PCC, hippocampus,

mPFC, DLPFC and parietal lobe activation was greater for hypothetical shocks, whereas TPJ, putamen, ACC and amygdala activation was greater for "real" shocks (Feldmen Hall et al., 2012). These results suggest that self interest (keeping money) interacts with social context - paying to avoid a bad outcome for yourself may be more motivating than paying to avoid a bad outcome for others. Finally, another study examined how the past experience of having an illness impacted hypothetical decisions about which illness would be preferable to have in the future, and if making those choices in turn reduced perceived aversiveness of the illnesses even further. They found that the interaction of past experience with predicted aversive outcome was related to increased signal in ACC and caudate (Sharot et al., 2010). Illnesses were perceived as less aversive if they were familiar, or if subjects chose that illness during a forced choice decision task, suggesting that both previous experience and choice affect aversiveness judgements as well as related brain activation. These studies raise important questions about whether behavior during hypothetical aversive choice differs from real aversive choice. Study 2 expands upon the findings of Sharot et al. (2010) by including similar stimuli (illnesses) and quantifying familiarity effects on perceived aversiveness, while also customizing aversive stimulus categories to participant's perceptions of severity. In contrast to Kang & Camerer (2013) and Feldmen Hall et al. (2012), we do not manipulate social context but instead focus on how individual differences in risk attitudes and familiarity with stimuli affect decisions for aversive outcomes. Additionally, we address how an approach versus avoidance context influences decisions for heterogenous, multidimensional aversive reinforcer types.

Risk attitudes can affect purchasing behavior as well as related brain activity particularly in choice contexts involving potentially negative consequences. Risk sensitivity tends to follow similar patterns cross-culturally, with most people being loss aversive after a gain and riskier/more loss seeking after a loss (Reiger et al., 2014). Thus, measuring risk attitudes is particularly important when studying loss contexts because the threat of loss itself tends to bias

behavior to be more risky. Additionally, the type of risks subjects are exposed to can lead to recruitment of context specific brain regions. For example in one study, subjects rated both their likelihood of buying items as well as the social risk of doing so while in the scanner (Yokoyama et al., 2014). While rating social risk of purchases explicitly, activation was observed in areas associated with social cognition (TPJ, mPFC). However when scores of a risk attitude questionnaire were used to predict activation while subjects rated likelihood of buying items, activation was seen in the insula. These results suggest that risk attitudes implicitly influence purchase decisions, independent of preference for purchase options. In order to quantify potential context dependent risk attitudes in Study 2, we use the DOSPERT (Domain Specific Risk Attitude Scale; Weber et al., 2002). The DOSPERT quantifies both risk behaviors and risk attitudes in domains such as financial, social, or health risks, which are used as behavioral covariates and predictors of brain activation to clarify how these factors influence behavior and brain activation during choices for aversive abstract reinforcers.

Using abstract reinforcers offers a unique advantage in answering questions in the aversive domain, because unlike stimuli such as electric shocks or monetary losses, they should not induce acute anticipatory anxiety, fear or dread because the scenarios are hypothetical. Abstract reinforcers also avoid the confound of the hedonic/sensory elements of pain, the logistical issues of implementing real world type losses in the lab (such as monetary penalties), and the lack of ecological validity in using stimuli such as electric shocks. Disentangling the valence of stimuli from the reinforcement they predict also removes potentially confounding goal motivations.

The key questions addressed by Study 2 involve whether in the abstract reinforcer context, avoiding a hypothetical negative outcome recruits the same brain regions (e.g. striatum, mPFC) as approaching an appetitive abstract reinforcer. In other words, does avoiding something bad involve activation in areas that also respond to reward? Is the subjective value of avoiding a

very bad thing higher than the value of avoiding something less bad, and are these value judgements represented in the brain via activation or deactivation in areas such as mPFC? We also aim to model interactions amongst both active and deactive regions using connectivity analysis in order to uniquely characterize the dynamics that lead to magnitude increases and decreases during decisions for aversivde stimuli.

Abstract Reinforcer Similarity

Abstract reinforcer stimuli provide a novel mechanism for studying the effects of context, value, and valence in the brain, but they also pose several challenges. While individualizing choice categories allows for a more customized way to study value, it also means that subjects in Studies 1 and 2 all make different and potentially unique choices. That is, even subjects who selected the same categories as high value or severity may view each choice differently - "hiking versus biking" might be a very difficult choice for one participant, and easy for another, each coding based on entirely different factors such as value, familiarity, etc. In many laboratory tasks, decision variables such as task difficulty, loss aversion, learning rate, etc. can be quantified exactly for all subjects because the number of choice dimensions that vary in the task are limited, such as probability and magnitude in a gambling task. With abstract reinforcers, there are essentially no bounds on the search space of individuals or the possible evaluative dimensions - that is, each choice could be made based on thousands of different factors. Despite these differences, Studies 1a+b and 2 demonstrate that meaningful group-level patterns of activation can be measured using this approach, and that these results are highly consistent with results from decision making studies using concrete stimuli. In Study 3, we aim to extend and clarify these results by having participants rate the similarity of abstract reinforcers instead of making preference judgements.

Using a similarity task allows us to test several key questions about the structure of abstract reinforcer choices. We can differentiate activation related to increases in value from

activation related to increases in similarity, to help clarify how similarity of stimuli informs preferences. For example, within a group of related stimuli, we may like one the best (hiking), and also prefer several others that are most similar to the favorite (running, walking). However, data from Study 1b in particular suggest that preferences for abstract reinforcers can involve more fine grained, idiosyncratic distinctions. For example, a subject who listed graphic novels as a top 3 preferred stimulus listed comic books as one of their 3 least favorites, despite the shared features of the two (see figure 3). Thus is it not clear if and how similarity and value interact during abstract reinforcer decisions. We aim to quantify the shared and distinct processes involved in preference judgements vs. similarity judgements at the level of brain activation in Study 3. We test whether similarity ratings involves activation in the same regions as preference judgements, specifically the striatum, mPFC, insula, and anterior cingulate. We also measure whether high vs. low similarity involves recruitment of the same or different regions than high vs. low value. Finally, because of the real-world validity of these stimuli in relationship to subject's preferences, we can measure how subjective value may influence or interact with behavior during similarity ratings.

Figure 3: Preference rankings and graphs during positive framing (left) and negative framing (right) for one subject during appetitive abstract reinforcer framing.

Research has demonstrated that value coding of stimuli occurs as a seemingly involuntary component process of choice even when it is task irrelevant, a phenomena called value based attentional capture (VBAC). Attention capture more generally is well characterized in psychophysics (Ruz & Lupianez, 2002). Research into this phenomenon suggests that the presentation of any irrelevant stimuli can capture attention, such when a symbol appears in the opposite corner as a fixation target, drawing eye gaze away from the target. The bottom-up account of this phenomenon compares this to when subjects involuntarily direct attention to a stimulus, such as turning one's head when one hears their name called. However, other studies suggest that attentional capture can be modulated by top down factors, such as limiting the presentation time of distractors (Kiss et al., 2012), changing the task set (Eimer & Kiss, 2008), controlling for visual features of the display (Leber & Egeth, 2006), or varying whether the cue is relevant to the task at hand (Lamy et al., 2014). Other results showing that working memory capacity strongly influences the ability to override attentional capture suggest that top-down processes can explain some VBAC response (Fukuda & Vogel, 2009). Other studies have found that the VBAC effect is amplified if the distractor stimuli have relevance in a different choice context. For example, color coded stimuli that previously predicted monetary wins drew attention away from target identification during a perceptual task (Anderson et al., 2011a). The degree of attentional capture by stimuli that previously predicted reward values was greater than the effect of visual salience (pop out effects of a brightly colored target; Anderson et al., 2011b).

Study 3 provides an ideal and reasonably naturalistic test case for bottom up vs. top down effects of VBAC. The key feature of this approach is that because subjects come into the study already having preferences for the stimuli used in the study, preferences do not need to be artificially induced through training (such as by learning that certain symbols predict monetary wins versus losses). All previous studies of VBAC for real world choices, to our knowledge,

require that subject's preferences be quantified in the lab prior to testing, which makes it impossible to determine how much of the VBAC effect is due to priming related to these procedures. For example, in Grueschow et al. (2015), participants made 672 preference judgements for movies 1-3 days prior to the VBAC experimental session. This training session may involve learning processes which artificially induce preference, e.g. via mere exposure effects. In Study 3 on the other hand, subject's preferences are based on their actual real world experiences, and they are not primed to think about value of stimuli via any pre-scan training. Instead, before scanning subjects select a category they prefer, and then in the scanner they rate similarity of stimuli in that category, a task that requires no explicit value coding. Only after the scan are subjects asked to rank order the stimuli by subjective value, by reporting the desired hours per week (dHPW) they would engage with their most preferred stimuli if time or money weren't an option (a measure analogous to willingness to pay). Then, dHPW scores can be used as a predictor of brain activation during the similarity rating task.

If value based attentional capture is an implicit, bottom up process, then dHPW scores should predict activation during similarity coding particularly in value related areas such as mPFC, similar to other recent studies (Greuschow et al., 2015). However, if dHPW scores are not related to activation during the perceptual task, this would suggest that VBAC for abstract reinforcer stimuli does not occur as a bottom up process. It would instead require a top-down mechanism, such as an effect of task set like assigning value ratings before completing the perceptual task. Using abstract reinforcer stimuli to test the top-down versus bottom up accounts of VBAC provides a better approximation of how VBAC affects real world decisions. That is, is the VBAC phenomena an artifact of the lab environment, a top-down effect induced by training? Or is it a bottom-up effect that occurs without explicit training?

In sum, Study 3 provides a framework for quantifying differences and similarities between preference judgements and similarity ratings for abstract reinforcers. We predict that due

to the inherent personal relevance of stimuli, similarity ratings will engage brain circuitry involved in value such as the striatum and mPFC. If this activation is truly value related then it is expected that dHPW ratings will predict activation in these areas as well, in line with previously observed results in Study 1b. Value related activation during similarity ratings would also further support a bottom-up account of value based attentional capture. Thus, Study 3 contributes important knowledge about how the brain codes abstract reinforcers, and provides a unique test case to determine how attention and value interact during decisions for preferences.

Summary

In sum, this proposal aims to characterize brain network response to decisions for abstract reinforcer stimuli. This response can then be characterized in comparison to well established dynamics that occur during concrete reinforcer decision making. The contexts used to test the potential overlap and differences between abstract and concrete reinforcer decisions include choice context, manipulated by choice prompt framing; valence, manipulated by using appetitive and aversive stimuli; value, manipulated by individualizing stimuli to subject's interests; and finally, using similarity ratings of abstract reinforcers to clarify brain coding of these stimuli and test hypotheses about value-based attentional capture.

Study 1a: Brain network response underlying decisions about abstract reinforcers

Decision making studies typically use tasks that involve concrete action-outcome contingencies, in which subjects do something and get something. No studies have addressed decision making involving abstract reinforcers, where there are no action-outcome contingencies and choices are entirely hypothetical. The present study examines these kinds of choices, as well as whether the same biases that exist for concrete reinforcer decisions, specifically framing effects, also apply during abstract reinforcer decisions. We use both General Linear Model as well as Bayes network connectivity analysis using the Independent Multi-sample Greedy Equivalence Search algorithm (IMaGES) to examine network response underlying choices for abstract reinforcers under positive and negative framing. We find for the first time that abstract reinforcer decisions activate the same network of brain regions as concrete reinforcer decisions, including the striatum, insula, anterior cingulate, and VMPFC, results that are further supported via comparison to a meta-analysis of decision making studies. Positive and negative framing activated different parts of this network, with stronger activation in VMPFC during negative framing and in DLPFC during positive, suggesting different decision making pathways depending on frame. These results were further clarified using connectivity analysis, which revealed stronger connections between anterior cingulate, insula, and accumbens during negative framing compared to positive. Taken together, these results suggest that not only do abstract reinforcer decisions rely on the same brain substrates as concrete reinforcers, but that the response underlying framing effects on abstract reinforcers also resemble those for concrete reinforcers, specifically increased limbic system connectivity during negative frames.

Introduction

The role of the corticostriatal circuit in motivation, reward seeking, and decision making has been well studied in the context of biologically relevant rewards. Several key dimensions typically define these choice scenarios. First, the *survival value* of stimuli used in studies ranges from the most pressing biological needs (food, shelter), to the means for acquiring those needs (money, social status), to choices more distant from survival such as choosing a vacation destination. Second, in many studies, the choices have *hedonic value,* in that the properties of the stimuli themselves activate hedonic drives (such as pictures of appetitive food invoking a salivary response). Third, in virtually all of these studies, there are clearly defined action-outcome contingencies in that the subject does something (e.g. presses a button to choose a gamble) and something happens (they win/lose). Thus, another dimension of the stimuli in these studies is that they involve *concrete reinforcers* - subjects expect and receive tangible experiences like winning money, food, viewing attractive faces, or listening to pleasant music. However, many of the decisions we make every day involve potential future states, not immediate outcomes - should we finish writing a paper now so that we have time to go for a bike ride later? What would be nicer, a bike ride or a hike? Daydreaming about or planning pleasurable experiences like hobbies involves an *abstract reinforcer*, in that it is hypothetical, not tangible. The brain response underlying these kinds of choices have not been addressed in the current literature.

The present study addresses this topic using a task in which participants make hypothetical choices with no expectation of outcomes, no actual outcomes, and no concrete reward for choices. This is done using a simple "which do you like more/which do you like less" decision prompt. We hypothesize that abstract reinforcers engage the same brain areas as concrete reinforcers, but the role these brain areas play will be different due to the different task demands for making decisions about abstract reinforcers. Specifically, as there are no explicit goals when choosing between abstract reinforcers, we expect that activation of the prefrontal cortex will

track the perceived or subjective value of the options presented rather than tracking goals as in concrete decision making. The anterior cingulate, rather than predicting outcomes of choices, since there are no outcomes, will instead direct attention and integrate affective feedback to guide choice behavior. The insula, instead of predicting risk, will integrate affective components such as how salient or arousing choice options are. Finally, the striatum will not track reward outcomes via a prediction error algorithm requiring terms for expectations and outcomes, but will instead track the perceived reward value or attractiveness of the choice options in order to guide decisions.

We also test whether the behavioral biases due to framing in concrete reinforcer scenarios also occur during decisions - using abstract reinforcers. We expect that negative frames will involve longer reaction times than positive frames in terms of behavior, as has been found in studies using concrete reinforcers (e.g. C. Alos-Ferrer et. al., 2012, Foo et. al., 2014). This longer RT for negative framing is thought to be related to increased negative affect, akin to loss aversion, although unlike in monetary contexts where subjects must evaluate offers for potentially bad gambles, here they are choosing which of two exemplars in a preferred category they must reject. This is thought to also induce choice conflict, particularly when being forced to choose between exemplars when both are highly preferred. Choice conflict plus increased negative affect are proposed to account for the longer reaction time in the negative condition, which should be reflected in engagement of brain areas implicated in processing aversive content, such as the ACC and insula. Differences in brain response during positive and negative framing will be addressed by examining both the magnitude of activation using GLM analysis as well as patterns of connectivity between a network of regions associated with decision making and reward, including VMPFC, dorsolateral prefrontal cortex (DLPFC), anterior cingulate (ACC), insula, caudate, and putamen.

Methods

Participants

Sixteen healthy adult participants (8 female, ages 19-60; mean age =25.47, SD=4.37 years) underwent functional MRI conducted at the Rutgers University Brain Imaging Center (RUBIC). Participants met standard MRI exclusion criteria (e.g.,. no metal implants, pregnancy, neurological disorders). One participant did not complete all experimental conditions and was excluded from the analysis. Three participants identified as white, 5 as asian, 2 as black, 3 as hispanic, 1 as pacific islander and 1 as other/multiracial. Participants were recruited from the Rutgers University Newark community through a department based subject recruitment system and word of mouth. Thirteen participants were undergraduates, 1 was a graduate student, and 1 was a staff member. Undergraduates were awarded course credit for participation. One participant reported taking medication, a low dose (5 mg) of the stimulant Adderall. Significantly higher doses of stimulants (20mg) have been found to modulate brain activity during attentional tasks (Tomasi et. al., 2011), but given the low attentional load of the present study and the low dose of the drug, this participant was included in the analysis after a review of their data showed reaction time within the range of the sample and patterns of brain activation consistent with the rest of the group. The same review process was applied to one left handed participant who was ultimately also included in the analysis. All participants gave informed consent to participate. The study was approved by the Rutgers Institutional Review Board.

Procedure

An Abstract Reinforcer Task (ART) was developed through behavioral piloting with an independent group of subjects (n=54) to determine an appropriate range of categories, exemplars within those categories, and to optimize task format. Participants selected a category that represented something they liked and were familiar with from the following options: sports, cultural activities, active lifestyle, historical periods, architectural styles, travel, countries, and

reading material (see APPENDIX for a list of all category exemplars). Categories chosen as preferred by subjects in this study included reading material (3 subjects), sports (4), countries (3), historical periods (4), and architecture styles (1).

In the scanner, participants made two-alternative forced choices between all possible combinations of category exemplars (i.e. "skiing vs. hiking"), once with the prompt "which do you like more" (positive framing) and once as "which do you like less" (negative framing), for a total of 132 trials. During each trial, the question prompt and two answer choices appeared on the screen. Participants had up to 8 seconds to respond via button box before the next set of answer choices appeared. Once the participant responded, the screen changed to a crosshair fixation for the remainder of the trial. This individually determined fixation period served as an inter-trial interval of variable length to help maximize statistical independence of trials. Stimuli were presented via an event related design paradigm (positive and negatively framed choices intermingled) with choice option order randomized for each participant. For a control condition, participants made positive or negatively framed size judgments (e.g. "which is bigger/smaller, pebble or boulder?") over a 30 second continuous block, and subjects also rested with their eyes open for a 30 second period. Stimuli were presented and responses recorded using PsychoPy (http://www.psychopy.org/).

Scanning Parameters

Functional imaging was conducted using a Siemens 3.0 Tesla Trio MRI scanner to acquire gradient echo T2*-weighted echo-planer (EPI) images with BOLD contrast. A 12 channel array coil was used due to increased signal detection in orbitofrontal regions. Each volume collected had 32 axial slices. 620 measurements were acquired in ascending contiguous order with a TR of 2s, for a total scan time of 21 minutes and 7 seconds. Imaging parameters included: field of view, 192 mm; slice thickness, 3mm; TR, 2s; TE, 30ms; flip angle, 90 degrees. Whole brain high resolution structural scans were acquired at 1 X 1 X 1 mm using an MP-RAGE pulse

sequence. An autoalign scout (AAS) sequence was used to orient the EPI images to the structural images, with a FOV of 260mm, TR of 3.15ms, and TE of 1.37ms.

fMRI General Linear Model

Analysis was performed using FMRIB's Software Library (www.fmrib.ox.ac.uk/fsl). Skull stripping was performed using BET (Brain Extraction Tool) and then registered using FLIRT (FSL's Linear Registration Tool), in which the BOLD functional data are registered to the MPRAGE anatomical scan and then to the MNI atlas image. FEAT (FSL's Expert Analysis Tool) was used for all GLM analysis with the following parameters for first level (individual scan) analysis: motion correction with MCFLIRT; 5 mm FWHM spatial smoothing, highpass filtering, and a second registration to the MNI atlas using 3 DOF. The following 7 regressors were used as predictors of brain activation: positive framing early (first 4s of 8s trial), positive framing late (last 4s of 8s trial), negative framing early, negative framing late, and size judgment baseline early and late. Crosshair fixation was not modeled.

Group level mixed effects analysis was conducted to compute average activation during conditions of interest. Head motion for the sample was minimal (<.5 mm; mean=.277, SD=.121) and thus movement was not used as a regressor. Demeaned reaction time values were entered as a regressor of no interest. Experimental conditions were compared against a crosshair rest baseline to generate average values at the individual subject level. For group averages, activation was tested for significant differences from zero using a t-test as implemented in FSL. T tests directly comparing conditions were performed as follows: positive framing > negative framing, negative framing > positive framing, using the FLAME 1+2 setting in FEAT which uses an iterative procedure to increase sensitivity in detecting significant task related activations. The early portion of the positive framing condition (first 4s of 8s trial) was tested against the late portion of the negative framing condition (last 4s), as decisions were made later in the trial during negative framing.

Connectivity

Connectivity analysis was performed to quantify how brain network response during decisions for abstract reinforcers is influenced by framing. While general linear model analysis addresses the level of response by various brain regions, it can not reveal how those brain regions interact. For example, if an area such as the insula is active at a similar magnitude in two task conditions, when those conditions are contrasted the insula will not appear in the statistical maps. However, non-significant differences in magnitude do not mean that the insula is not engaged in the two conditions. The insula may be activated to a similar level in both conditions, but may play very different roles in terms of the connectivity of activated brain regions.

Here, we use an Independent Multi-sample Greedy Equivalence Search (IMaGES), which operates on the standard assumption that causes precede effects temporally. Causality in this case refers to the ability of IMaGES, in conjunction with a distributional post processing step (LOFS), to identify directionality of connections with high probability (Ramsey, Hanson & Glymour, 2011). The algorithm starts with an empty graph and searches forward, one new connection at a time, until it finds the set of connections that optimally represents the entire group of subjects, interpolating any missing data. Finding oriented edges requires two steps, first the estimate of connectivity (provided by GES and individual Bayes Information Criteria (BIC) constraints on time series regressions); and second, a method that systematically investigates conditional dependence/independence and estimates orientation. The algorithm searches with the restriction of finding only Markov equivalence classes of directed acyclic graphs, and without the option of varying time lags, given that doing so did not improve accuracy in simulation (see Ramsey et. al. 2010, 2011 for a more in depth methods discussion). The process is penalized to prevent overfitting using the Bayes Information Criterion (Schwarz, 1978): $-2\ln(ML) + k \ln(n)$, where ML is the maximum likelihood estimate, k is the dimension of the model (the number of directed edges plus the number of variables), and n is the sample size (number of participants). As the

reliability of the IMaGES algorithm was found to be higher when using the LOFS search post filter, which determines and assigns the "dominant" direction of the edge (removing bidirectional edges), the current analysis was run using this option and thus only unidirectional edges were returned.(Ramsey et. al. 2011).

ROIs were chosen based on activation during GLM analysis as well as previous literature and included VMPFC, DLPFC, ACC, caudate, putamen, and insula. Binary masks were created for bilateral regions of interest using FSL view and the Harvard-Oxford anatomical atlas, in which the atlas defined ROIs were converted into masks. Average time series for each subject were extracted from these ROIs using FSL's meanTS module. Time courses of interest were arranged into a matrix for each subject, with the ROIs as columns and each row representing a single time point. These files were then input into the IMaGES workflow in Tetrad. Graph selection was conducted by choosing the graph with the highest BIC score. Edge (connection) weights were exported from Tetrad into LibreOffice Calc (https://www.libreoffice.org/). T statistics were averaged across the group, and were used instead of raw coefficient values because they take into account standard error. The TDIST function was used to calculate significance values.

Results

Behavioral

RT was significantly longer for negative framing (M= 2.72, SE=.13) compared to positive (M =2.47s, SE= .13; t = -5.4179, df = 14, p-value = .000091). Negative framing had significantly longer RT than the negative size comparison baseline (M=2.18, SE=.13, t = 9.4099, df = 14, p-value = 0.000002). Positive framing also had significantly longer RT than the positive size comparison baseline (M=2.1, SE=.09, t = 5.154, df =14, p-value = 0.00015).

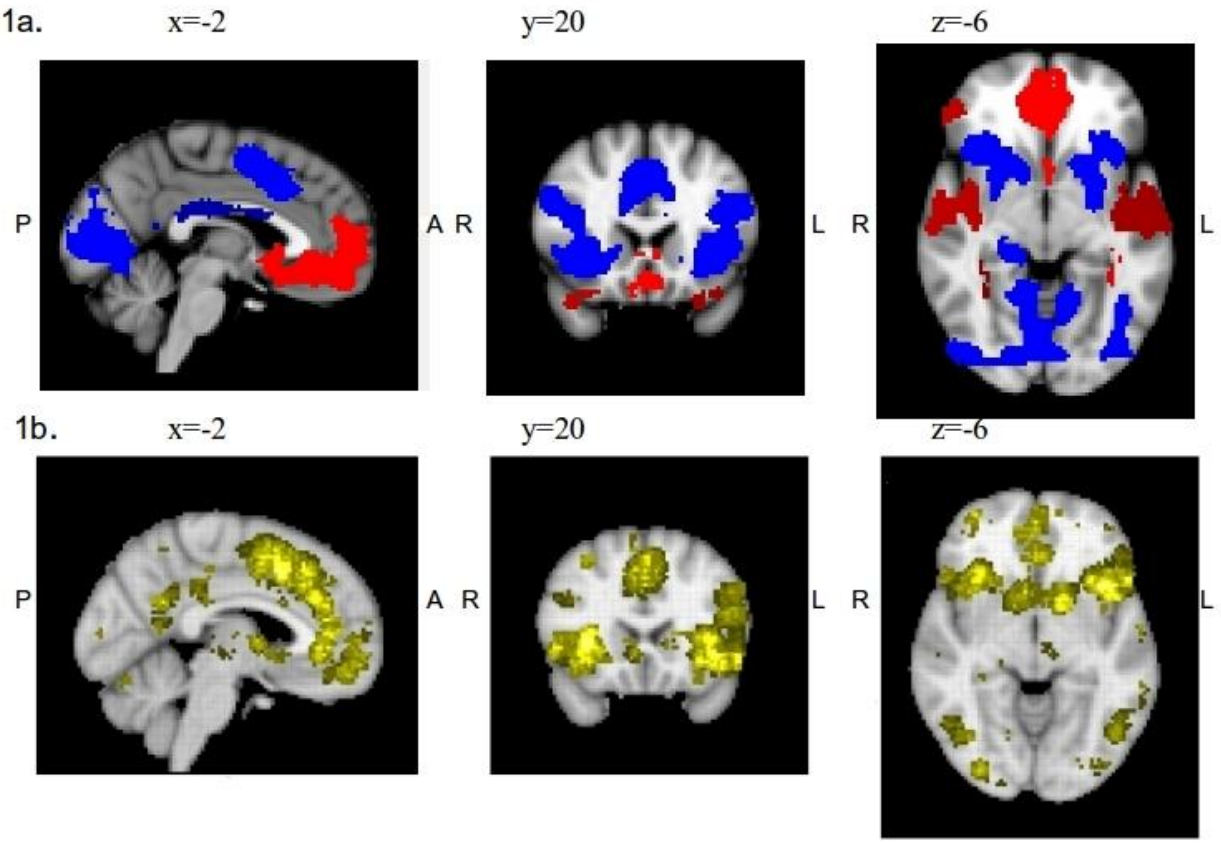


Figure 2.1a+b. Comparison of Neurosynth meta-analysis search to study results. Both slices pictured at the coordinates x=-2, y=20, z=-6. A) Activation greater in positive framing>negative framing (blue) and negative framing>positive framing (red). Results are shown at a cluster threshold of $p<.05$. B) Results of a Neurosynth search for brain areas active in studies about decision making.

GLM

Activation for the group average of the positive framing condition was seen in pre and post central gyrus, lateral occipital cortex, brainstem, bilateral middle frontal gyrus, anterior and posterior cingulate, paracingulate, ventral striatum, left thalamus, insula, lingual gyrus, dorsolateral prefrontal cortex (DLPFC), and precuneus. In the contrast of positive framing>negative framing, the same areas were active as in the group average of the positive

condition with the exception of additional activation observed bilaterally in the thalamus, inferior temporal gyrus, and right caudate (Figure 2.1a, in blue).

In the group average of the negative framing condition, activation was observed in bilateral postcentral gyrus, posterior cingulate, bilateral thalamus, bilateral ventral striatum, left dorsal striatum bilateral insula, bilateral central opercular cortex, bilateral precuneus, bilateral superior temporal gyrus, and bilateral temporal pole.

In the negative framing>positive framing contrast, areas with significant activation included the bilateral lateral occipital cortex, left angular gyrus, right supramarginal gyrus, left dorsomedial prefrontal cortex, left precentral gyrus, right postcentral gyrus, bilateral inferior and middle temporal gyrus, anterior and posterior cingulate, bilateral superior temporal gyrus, bilateral thalamus, bilateral amygdala, bilateral orbitofrontal cortex, ventromedial prefrontal cortex (VMPFC), left thalamus, bilateral ventral striatum, insula, and precuneus (Figure 2.1a, in red).

Neurosynth Comparison

Given the novel nature of the task format, we compared results from our GLM analysis to a meta-analysis of decision making studies using Neurosynth. Neurosynth is an actively maintained, automated meta-analysis tool which allows the user to search their database of studies (numbering 5,809 at time of analysis) for common patterns of activation by using targeted search terms such as "working memory" or "emotion regulation." For this analysis, the search terms used were "decision," "decisions," and "decision making" (www.neurosynth.org). The activation patterns retrieved from Neurosynth are pictured in Figure 2.1b, and results from both conditions of our analysis are pictured overlayed together on a structural image, with results for the negative framing condition in red and positive in blue (in both images the slice coordinates are x=-2, y=20, z=-6). The pattern of activation during our experimental conditions closely matches the pattern identified using Neurosynth.

Connectivity

A connectivity graph was calculated for the following regions: DLPFC, VMPFC, insula, anterior cingulate, caudate and accumbens for both positive and negative framing. The models with the highest Bayes Information Criteria scores were selected (BIC= -9279.8 for negative framing, and BIC=-9226.8 for positive). The Harvard-Oxford atlas regions used as masks to extract timeseries are shown in Figure 2.2. These regions were chosen based on their activation in the GLM analysis and the extensive literature on their roles in decision making. Since the function of these areas is computationally well defined in concrete decision making contexts, this allows for more informed predictions for how these areas might function in an abstract contexts (e.g., that rather than predicting risk, the insula will integrate affective components such as how salient or arousing abstract reinforcers are). Additionally, limiting the areas included in the graph to six helps constrain the potential number of connections found in order to ensure graphs are interpretable (e.g., a 6 node graph with 7 connections is easier to parse than a 10, 15, or 20 node graph likely to have dozens of connections).

2.

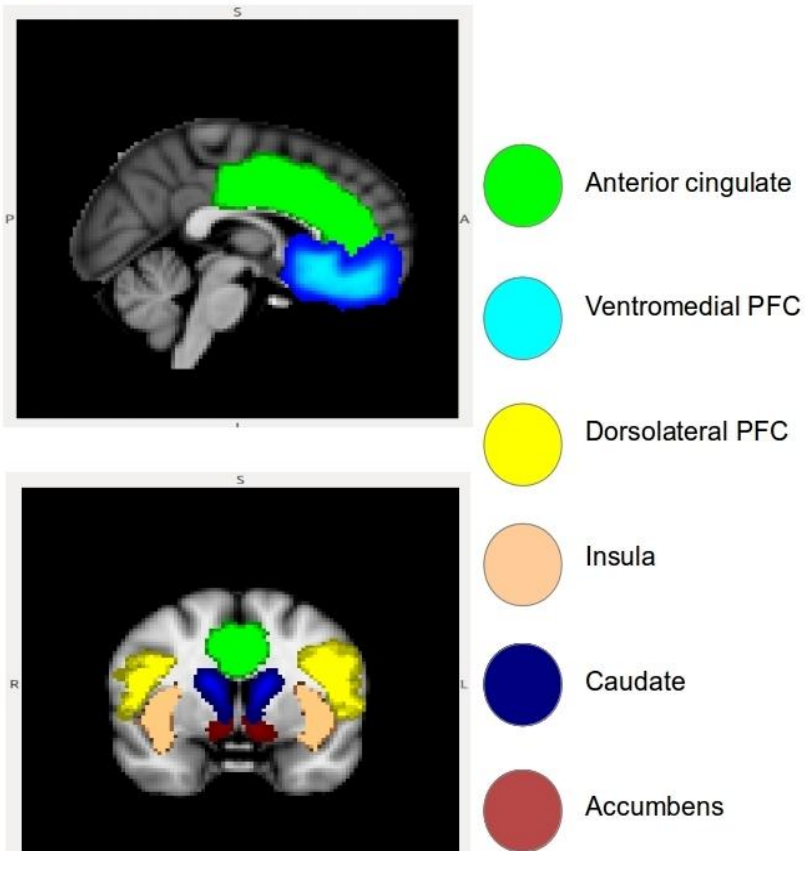


Figure 2.2. Masks used to extract timeseries data from regions included in connectivity analysis. Masks were derived from the Harvard-Oxford anatomical atlas and exported using FSLview.

Connections common to both conditions included: DLPFC to ACC; DLPFC to accumbens; accumbens to VMPFC; and caudate to ACC (shown in yellow in Figure 2.3; note that regions are not displayed in actual anatomical locations; refer to figure 2.2 for anatomically defined masks). These connections did not differ significantly between conditions in direction or in coefficient strength. The direction of connections between ACC and insula, and insula and accumbens, varied by condition as well as coefficient strength (shown in blue, Figure 2.3). The connection from ACC to insula was significantly stronger in the negative framing condition ($t = -3.4924$, $df = 14$, p-value = 0.0036) as was the connection between DLPFC and ACC ($t = 3.6677$, $df = 14$, p-value = 0.0025). Connectivity strength between the caudate and ACC was also stronger in the negative condition ($t = 2.3113$, $df = 14$, p-value = 0.03656). Finally, the connection between the insula and accumbens differed between conditions in several ways: it had a negative coefficient during negative framing, indicating downregulation of the accumbens by the insula; the direction of the connection changed from accumbens-insula in positive framing to insula-ACC in negative; and the connection between these regions was significantly stronger in negative framing ($t = 6.425$, $df = 14$, p-value = .0000016). Using a Bonferroni multiple comparisons

correction, three of the four edge weight differences survive correction with an error rate of 2.77%, with the caudate-ACC edge not surviving correction.

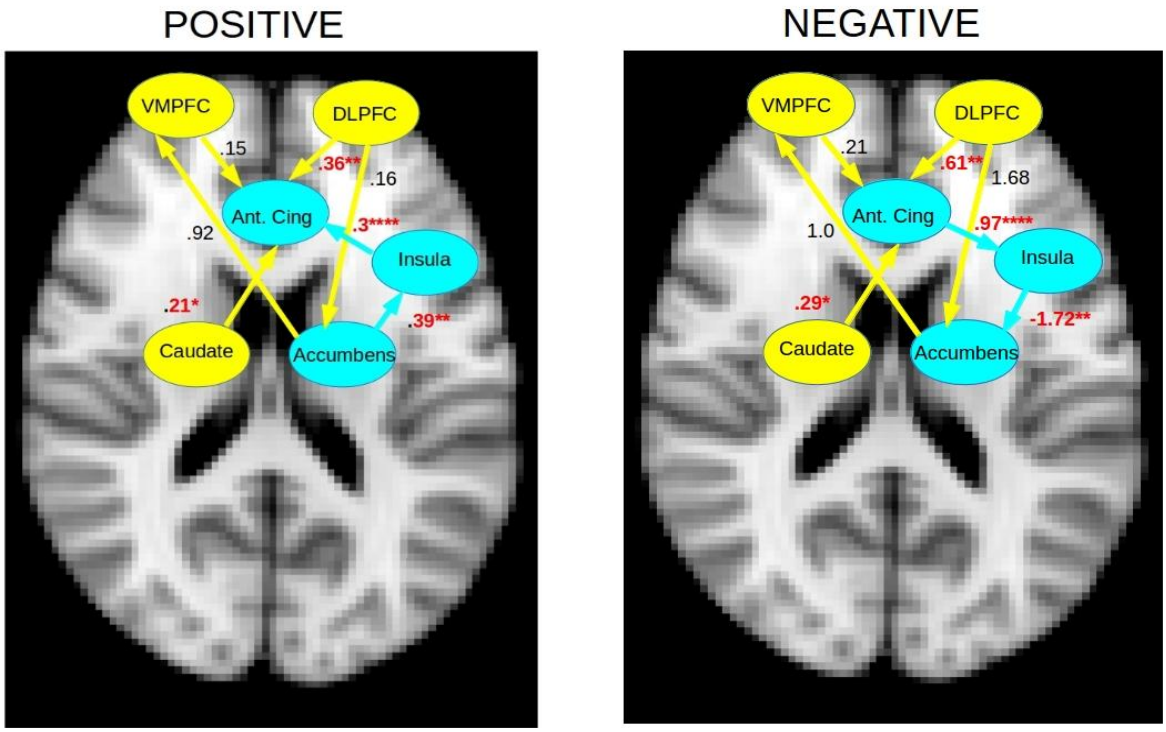


Figure 2.3. Connectivity analysis of positive and negative framing. Regions are not pictured in actual anatomical locations. Connections (edges) common to both conditions are in yellow and those that vary between conditions are highlighted in blue. Numerical values next to edges are coefficient values. * =p<.05, **=p<.005, ****= p<.00001

Discussion

Results suggest that brain regions are recruited differently to support choice evaluation and selection during positive versus negative framing of decisions about abstract reinforcers. Concordance between the results of this study and a Neurosynth meta-analytical search for decision making studies suggest that decisions about abstract reinforcers activate the same network as concrete decision making, including areas of striatum and VMPFC previously associated with concrete reinforcers only. However, different parts of the network are active in each respective framing condition, with more activation in VMPFC during negative framing and DLPFC in positive framing when compared using GLM analysis. This pattern is consistent with a pathway implicated in cognitive evaluation and general motivation, in which coordination between striatum and DLPFC is responsible for orientation to stimuli (Bromberg-Martin et. al., 2010). As it was hypothesized that positive framing involves a more simple, cognitively efficient selection process it is consistent with our hypothesis that the dorsolateral region of PFC activated is more associated with attention and orientation. In contrast, in the negative condition there is more activation in VMPFC. This pathway has been implicated in a more evaluative mechanism, consistent with the need to compare two options directly in order to reject one of them, even though both might be preferred (Bromberg-Martin, 2010). The VMPFC has also been shown to be crucial for sensitivity to negative consequences in lesion studies (Bechara et. al., 2000). Thus, perceived loss during negative framing from being forced to reject an option (one that may be well liked, just not as much as the chosen item) may underly the differences in brain activation in negative compared to positive framing.

Several related studies involving monetary and food rewards framed as losses or gains found activation in similar areas, but with varying patterns. In several studies, the negative/loss frame condition was related to decreased activation in prefrontal regions compared to either neutral or gain based conditions (e.g. Brieter et. al., 2001; Tom et. al., 2006). In another recent study, activation in VMPFC in particular was related to changes in value regardless of whether

the frame was negative or positive (Foo et al., 2014). In our results, VMPFC is more active during negative framing. During negative framing, subjects must actively evaluate attributes of each stimulus with the goal of determining which stimulus is less desirable, which may invoke valuation processes to a greater degree than when simply selecting the preferred option as in positive framing. Additionally, studies that have found reduced or deactivation of VMPFC during negative frames have had actual losses or aversive stimuli during these conditions. The present study only presented stimuli that ranged from neutral (non-preferred exemplars within a preferred category) to high value (most strongly preferred exemplars). Future studies could include aversive exemplars to clarify this effect.

Connectivity analysis revealed a common network structure in both conditions involving connections between the DLPFC, accumbens, caudate and VMPFC. Differences were observed in connections between ACC, insula, and accumbens. Connections to the ACC were stronger in negative framing for the caudate-ACC, accumbens-insula, and DLPFC-ACC connections. Although the caudate-ACC connection strength change did not survive multiple comparison correction, and thus should be interpreted with some caution, the overall pattern of stronger ACC connectivity in negative framing is consistent with the literature. Specifically, previous research has supported the idea that the ACC serves as a "hub" integrating decision variables such as perceived value, potential enjoyment, cognitive effort, and affective reactions (Chaudhry et. al, 2009; Etkin et. al., 2011; Shackman et. al., 2011; Gasquoine 2011, Shenhav et. al., 2013). Given the highly multidimensional nature of abstract decisions, it follows that an area such as ACC would be needed to integrate information and that this need would be especially great in negative framing, in which the decision process might include greater choice conflict and more negative affective components. The stronger connectivity between ACC and insula and insula and accumbens in negative framing compared to positive is also consistent with our hypothesis that choosing which exemplar to reject involves a stronger affective reaction than simply selecting

which exemplar is preferred. A potential explanation for this is that the ACC signal integrates decision variables communicated through DLPFC and VMPFC in order to help resolve choice conflict. The ACC is connected to the insula with a significantly stronger coefficient in negative framing compared to positive, which may indicate processing of affective content given the role of the insula in signaling risk and potential loss (see meta-analysis by Sescousse et. al., 2013). In addition, the connection between insula and accumbens has a negative coefficient, indicating down regulation - as insula activation increases in magnitude, the accumbens decreases in a related manner. This may indicate that the loss related information integrated in the insula interacts with the subjective reward signal being calculated in the accumbens. Once this reward signal is processed in the accumbens, it may then be communicated to VMPFC so this information can be integrated in order to guide action selection (Grabenhorst & Rolls, 2011). Connectivity between VMPFC and other areas did not differ between conditions, despite differences in VMPFC magnitude observed via GLM analysis. It is important to note that connectivity is measuring patterns of activation irrespective of magnitude, such that high magnitude activation does not necessarily imply strong connectivity and low magnitude activation may not be related to low connectivity strength (e.g. an area active at low magnitude as measured by GLM might still be strongly connected to the network as measured by connectivity, and vice versa). Thus the similar connectivity patterns of VMPFC in positive and negative framing suggests that in terms of it's interactivity with other parts of the network, it plays a similar role in both conditions . It is possible that the magnitude differences detected via GLM reflect instead a signal coding difference (e.g. increase in neuronal firing proportional to increases in perceived value; Tremblay & Schultz, 1999, Bouret & Richmond, 2010). Since connectivity and GLM measure different things, it seems likely that here these methods are capturing two potentially different aspects of VMPFC activation – network interactions and signal coding.

These results help to clarify the differences in the mechanism underlying decision processes for abstract reinforcers that are positively or negatively framed. In concrete reinforcer tasks in which framing is linked to a perceived outcome such as the value of a gamble, there are often probabilities that subjects must calculate, a process affected by factors such as subject numeracy (Peters & Levin, 2008). As we have shown, decisions involving abstract reinforcers are affected by the framing effect much as those involving concrete reinforcers. It should be noted that in the abstract case, the effect on both behavior and brain relies solely on a subjective perception of the choice which is itself easily influenced by the framing of the type of decision to be made. We conclude that choosing between abstract reinforcers requires coordination of the same network that supports decisions between concrete reinforcers, including areas implicated in affect and decision making.

Study 1b: Reinforcement value in the eye of the beholder? mPFC is modulated by individual preference for abstract reinforcers

A large body of recent research has supported a role for the medial prefrontal cortex (mPFC) in tracking subjective or perceived value (Grabenhorst & Rolls, 2011; Roy et al., 2012; Levy & Glimcher, 2012; Bartra et al., 2013; Clithero & Rangel, 2014). Several key features are shared by many of these studies. First, subjects often make value judgments about constrained, predetermined sets of low dimensionality stimuli such as food or money. Subjects are usually rewarded for their choices, such as by having one gamble randomly selected for an actual payout - a *concrete* reinforcement. While some studies have investigated real versus mentally simulated rewards (Bray et al., 2010 & Miyapuram et al., 2012), no studies have investigated mPFC value coding when choices are unrelated to any real or simulated outcome. Second, subjects typically choose from the same sets of stimuli which are assumed to be assessed on a common scale. That is, nearly all subjects are likely to rank $5 as more preferable than $3, or cake as preferable to broccoli. Far fewer studies investigate value computations when subjects choose from stimuli from entirely different categories with highly multidimensional features (e.g., Gross et al., 2014). In such contexts no objective common scale exists, and value is based on individual, ideographic assessments of stimuli. No studies to our knowledge have investigated whether mPFC codes value when choices are both hypothetical and lack a common evaluative scale. Here, we utilize a task context involving *abstract* reinforcers - hypothetical choices for options such as preferred hobbies or travel destinations, in which there are no perceived or actual action-outcome contingencies - to establish whether mPFC codes value in this case.

The aim of this study is to clarify whether mPFC responds to value specifically in an abstract choice context involving no action-outcome contingencies, goal motivation, overlapping

sensory features or inherent salience of stimuli, while controlling for familiarity and self-referential processing. We use an Abstract Reinforcer Task (Mills-Finnerty et. al., 2014), in which subjects make hypothetical choices amongst individualized sets of highly multidimensional stimuli, to address the role of mPFC in this decision context. Specifically, if mPFC responds in the absence of any actual or imagined decision outcomes, this would suggest that mPFC can code value in order to motivate choice behavior independent of achieving a specific goal (e.g., choosing for the sake of choosing, rather than doing so strategically to gain reward or avoid loss). Further indication that mPFC responds to content value rather than outcome value in the abstract context would be shown if mPFC activation covaried with choice value such that increased perceived value would elicit increased activation.

Methods

Participants

Sixteen healthy adult participants (8 female, ages 19-60; mean age =25.47, SD=4.37 years) underwent functional MRI conducted at the Rutgers University Brain Imaging Center (RUBIC). Participants met standard MRI exclusion criteria (e.g.,. no metal implants, pregnancy, neurological disorders). One participant did not complete all experimental conditions and was excluded from the analysis. Three participants identified as white, five as asian, two as black, three as hispanic, one as pacific islander and one as other/multiracial. Participants were recruited from the Rutgers University Newark community through a department based subject recruitment system and word of mouth. Thirteen participants were undergraduates, one was a graduate student, and one was a staff member. Undergraduates were awarded course credit for participation. One participant reported taking medication, a low dose (five mg) of the stimulant Adderall. Significantly higher doses of stimulants (20mg) have been found to modulate brain activity during attentional tasks (Tomasi et. al., 2011), but given the low attentional load of the present study and the low dose of the drug, this participant was included in the analysis after a

review of their data showed reaction time within the range of the sample and patterns of brain activation consistent with the rest of the group. The same review process was applied to one left handed participant who was ultimately also included in the analysis. All participants gave informed consent to participate. The study was approved by the Rutgers Institutional Review Board.

Procedure

The Abstract Reinforcer Task (Mills-Finnerty et al., 2014) was used to create an individualized choice paradigm for each subject. Participants selected a category that represented something they liked and were familiar with (high value category). We used both familiarity and liking to guide selection of the high value category to ensure that participants had measurable involvement with preferred stimuli, to generate familiarity ratings for use as a control measure. They also chose a category they didn't like and/or were less familiar with (low value category) from among the same general set. All subjects were able to identify a category they did not like except for one subject, who chose their low value category (sports) on the basis of being less familiar with it. The eight categories included: sports, cultural activities, active lifestyle, historical periods, architectural styles, travel, countries, and reading material. Categories chosen as high value by subjects in this study included reading material (three subjects), sports (four), countries (three), historical periods (four), and architecture styles (one). Categories chosen as low value included travel (two), sports (four), architecture (three), reading (one), active lifestyle (two), historical periods (two), and cultural activities (one).

In the scanner, participants made two-alternative forced choices between all possible combinations of category exemplars (i.e. "skiing vs. hiking"), under two different choice prompt conditions, for a total of 132 trials per condition. During each trial, a choice prompt (either "which do you like more" or "which do you like less") and two answer choices appeared on the screen. Participants had up to 8 seconds to respond via button box before the next set of answer

choices appeared. Once the participant responded, the screen changed to a crosshair fixation for the remainder of the trial. This individually determined fixation period served as an inter-trial interval of variable length to help maximize statistical independence of trials. Stimuli were presented with choice option order randomized for each participant. Subjects also rested with their eyes open for a 30 second period. Stimuli were presented and responses recorded using PsychoPy (http://www.psychopy.org/).

After scanning, participants filled out a questionnaire in which they answered questions about their involvement with their top three preferred high value category exemplars. For example, a subject who chose active lifestyle as their high value category might answer questions about yoga, running, and hiking. Participants were only asked to rate their top three exemplars because analysis of choice behavior revealed that preferences appeared to be stable for those choices over time, but that preferences for exemplars 4-9 in the participant's preference ranking were unstable and choices for options ranked #10-12 returned to stability (Supplemental Figure 2). Choice stability was measured using the graph-opt algorithm implemented in R's igraph network analysis package (http://cran.r-project.org/web/packages/igraph/index.html).

After scanning, participants were asked "how often do you engage in this activity every week (i.e. reading this material, traveling in this manner, playing or watching this sport, studying this style of architecture or historical period)" to measure actual hours per week (aHPW) and "if time or money weren't an obstacle, how much would you like to engage in this activity each week?" for desired HPW (dHPW). Finally, in order to measure the relationship between participant's self concept and preferred stimuli, they were also asked "How much do you feel like engaging in this choice activity is part of your identity (i.e. you would not be the "same person" if you were unable to read your favorite reading material, do yoga or your favorite activity, play or watch your favorite sport, engage in your favorite form of travel, etc.)?" Participants provided a rating from 1-4 as follows: 1=Not a part of my identity at all 2=slightly part of my identity 3=

moderately part of my identity 4=significant part of my identity.

Scanning Parameters

Functional imaging was conducted using a Siemens 3.0 Tesla Trio MRI scanner to acquire gradient echo T2*-weighted echo-planer (EPI) images with BOLD contrast. A 12 channel array coil was used due to increased signal detection in orbitofrontal regions. Each volume collected had 32 axial slices. 620 measurements were acquired in ascending contiguous order with a TR of 2s, for a total scan time of 21 minutes and 7 seconds. Imaging parameters included: field of view, 192 mm; slice thickness, 3mm; TR, 2s; TE, 30ms; flip angle, 90 degrees. Whole brain high resolution structural scans were acquired at 1 X 1 X 1 mm using an MP-RAGE pulse sequence. An autoalign scout (AAS) sequence was used to orient the EPI images to the structural images, with a FOV of 260mm, TR of 3.15ms, and TE of 1.37ms.

Behavioral Analysis

Stimuli were presented and behavioral responses recorded using Psychopy (http://www.psychopy.org/). Reaction time averages were calculated using LibreOffice Calc (http://www.libreoffice.org/). Statistics such as T tests on reaction time data were performed using R (http://www.r-project.org/).

fMRI Analysis

Analysis was performed using FMRIB's Software Library (www.fmrib.ox.ac.uk/fsl). Data for all subjects was skull stripped using BET (Brain Extraction Tool) and then registered using FLIRT (FSL's Linear Registration Tool), in which the BOLD functional data was registered to the MPRAGE anatomical scan and then to the MNI atlas image. FEAT (FSL's Expert Analysis Tool) was used for all GLM analysis with the following parameters for first level (individual scan) analysis: motion correction with MCFLIRT; 5 mm FWHM spatial smoothing, highpass filtering, and a second registration to the MNI atlas using 3 DOF.

Subjects made high value and low value choices during two separate scan runs, both totalling 21 minutes and 7 seconds. These separate runs were then processed using identical analysis streams, as follows: at the individual subject level, trials were separated based on choice prompt and trial timing (early or late). Results related to differences in choice prompts are reported elsewhere (Mills-Finnerty et. al, 2014). Trial timing was divided into early (first 4s of 8s trial) or late (last 4s) in order to account for reaction time differences related to choice prompt. In total, 7 regressors were used: positive early and late, negative early and late, size judgement control condition early and late, plus reaction time as a regressor of no interest. Then, the first 4s of the positive trials and last 4s of the negative trials were collapsed together in a second level analysis to modelled the average activation for each subject during both high and low value choices. Next, the individual files representing each subjects average activation during high and low value choices were averaged together using a mixed effects design. This produced a high value group average and a low value group average. The per-subject activation files were also used as input into a model that compared the high value condition directly to the low value condition using a T test and a mixed effects design.

Additional analyses included demeaned desired and actual HPW scores as well as self-concept ratings for the choices ranked #1, #2 and #3 by subjects, used as a predictor of brain activation in the high value condition. Reaction time was included as a covariate of no interest in these analyses and did not predict activation or alter results. This brain activation related to desired HPW was used as a contrast mask of the high value > low value analysis, to find areas of activation common to both analyses.

Results

Behavioral

Ratings of desired HPW (dHPW) for participants #1 preferred stimulus were negatively correlated with reaction time in the high value condition (Figure 3.1; $t = -3.3675$, $df = 13$, p-value

= 0.005). Both RT and dHPW were log transformed to correct for skew. Ratings of actual HPW (aHPW) were not correlated with reaction time or any other measure. Ratings of the relationship between the #1 preferred stimulus and self concept (M=2.87, SD= .83) were also not related to reaction time or other measures. The different between aHPW and dHPW was significant for participants top 3 choices (Figure 3.2; #1 aHPW vs. dHPW - t = -2.469, df = 14, p-value = 0.03; #2 aHPW vs. dHPW: t = -2.1229, df = 14, p-value = 0.05; #3 aHPW vs. dHPW -M=3.29, SD=3.27, t = -2.3924, df = 14, p-value = 0.03). Data on aHPW and dHPW for the #3 stimulus were missing for one subject and so an interpolated average value was used.

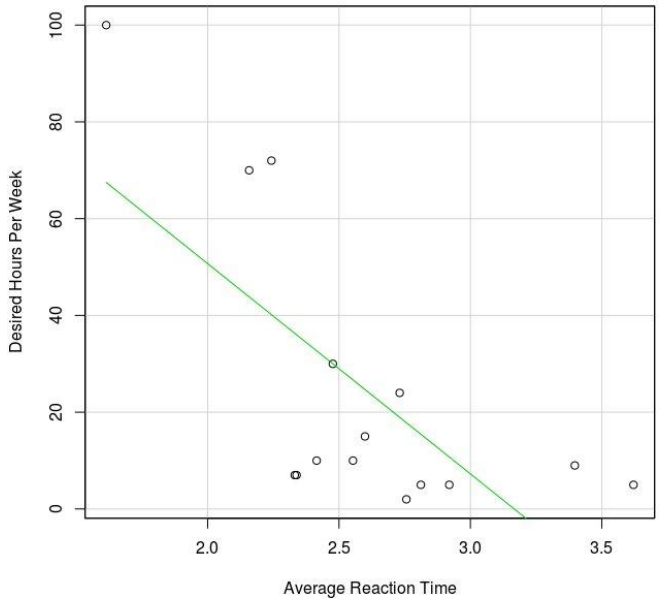


Figure 3.1. Correlation between reaction time values and dHPW.

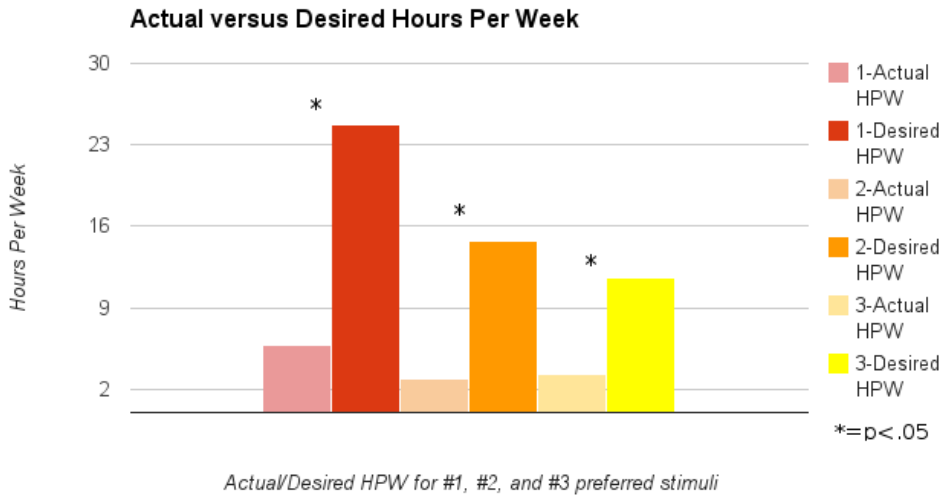


Figure 3.2. Differences between actual HPW (light bars) and desired HPW (brighter bars).

fMRI

Activation was observed in the group average of the high value condition in bilateral parietal and central operculum cortex, bilateral insula, and right superior temporal gyrus. All reported results were cluster thresholded at $p<.05$ unless otherwise noted.

In the low value condition group average, activation was observed in the anterior cingulate, juxtapositional lobule cortex, left pre and postcentral gyrus, left supramarginal gyrus, left lateral occipital gyrus, right lingual gyrus, intercalcarine cortex, left putamen and insula. In the contrast of high value> low value, activation was observed in mPFC and anterior cingulate (Figure 3.3).

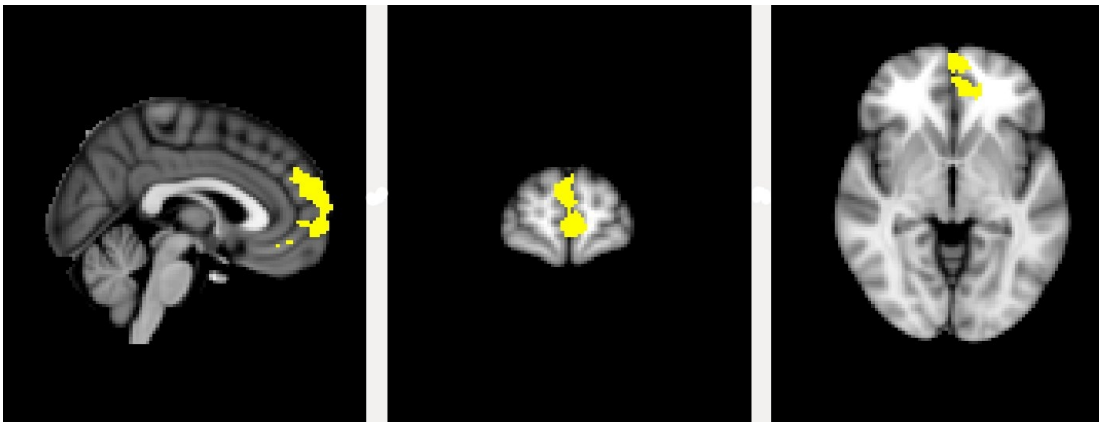


Figure 3.3. Contrast of high value > low value activation.

Activation for the group average of high value with desired HPW as a predictor was observed in mPFC (not pictured). No activation was observed above threshold when actual HPW or self concept ratings for the #1, #2, or #3 preferred stimuli were used as a predictor. The conjunction analysis of the activation predicted by desired HPW and the activation from the high value >low value contrast is shown in Figure 3.4.

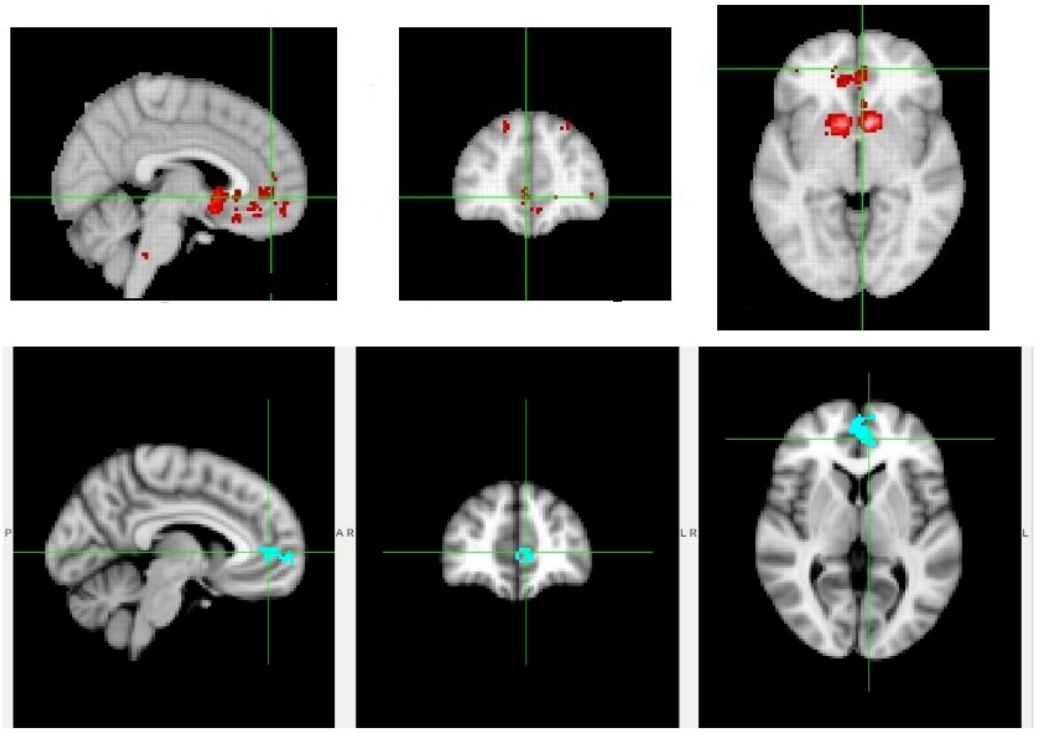


Figure 3.4. Top (red), coordinates related to value in a Neurosynth meta analysis. Bottom (blue):

conjunction analysis of the activation predicted by desired HPW and the activation from the high value >low value contrast.

Neurosynth Comparison

We compared results from our GLM analysis to a meta-analysis using Neurosynth to confirm that the area activated in our study is an area of prefrontal cortex associated with value. Neurosynth is an actively maintained, automated meta-analysis tool which allows the user to search their database of studies (numbering 9,721 at time of analysis) for common patterns of activation by using targeted search terms such as "working memory" or "emotion regulation." For this analysis, the search term used was "value" and returned coordinates generated from 288 studies (www.neurosynth.org). The activation patterns retrieved from Neurosynth are pictured in Figure 3.4 (top), and results from our analysis are pictured below. The cluster of activation generated from the conjunction analysis of the activation predicted by desired HPW and the activation from the high value >low value contrast encompasses the same coordinate range as the Neurosynth search results for value.

Discussion

We examined brain activation in subjects making choices about abstract reinforcers in categories individualized to be perceived by them as higher or lower in value. We find that when activation for the high value condition was compared against that for the low value condition, a cluster of activation was observed in mPFC. These results demonstrate for the first time that abstract reinforcer value is coded by mPFC in a similar manner as concrete reinforcers. This activation was also concordant with voxels responsive to value across a meta-analysis of 288 studies conducted using Neurosynth (www.neurosynth.org). These results suggests that mPFC is sensitive to increases in perceived stimulus value, independent of the presence of a choice outcome or goal motivation such as simulating the experience of a concrete reward. We

confirmed that mPFC response is specific to value by quantifying subjective value as desired hours per week (dHPW) subjects would engage with preferred stimuli if time or money weren't an option, and using these values as a predictor of brain activation. We found an overlapping set of voxels in mPFC are active both during the contrast of high value vs. low value conditions and when dHPW was used as a predictor. Actual HPW and ratings of the self-relatedness of stimuli did not predict any activation above threshold, suggesting that neither familiarity or self-relevant processing account for the observed mPFC response.

Our stimuli also inherently control for differences in stimulus salience and sensory processes. In this study, subjects largely chose different categories as their high value and low value sets - 11 out of 15 subjects chose different combinations of high and low value categories, and the four subjects who chose the same sets of categories had different preferred stimuli within those categories. Given that the stimuli within categories are highly variable in their stimulus properties, ranging from hiking to traveling to China to reading a novel, it seems unlikely that the greater activation in mPFC during the high value category is a reflection of specific stimulus attributes. These stimuli also vary in terms of the potential evaluative dimensions - assigning value to sports, for example, means taking into account entirely different stimulus properties than assigning value to travel destinations - making coding based on unidimensional sensory features unlikely. Our results show activation in a similar location as another study that used multidimensional and heterogeneous rewards but analysed value coding using multi-voxel pattern analysis (Gross et. al., 2014). These results suggest that this sub-region of mPFC may have some functional specificity for coding value of multidimensional rewards. Although more studies are needed to clarify whether this is the case, the issue of whether information content varies across mPFC regions that code value is an important emerging topic in decision neuroscience (Clithero & Rangel, 2014).

We also measured whether familiarity with stimuli was driving the signal observed in mPFC. We used subjects' actual HPW, reflecting current participation in preferred activities, and found it was not related to any brain activation above threshold. Actual HPW varied across participants; for example, participants who preferred reading reported engaging in more actual hours per week of the activity than subjects who preferred sports. If these familiarity differences were driving brain response, they should predict differences in brain activation; however, actual hours per week did not predict any brain activation above threshold. The positive relationship between desired HPW and mPFC activation suggests a dissociation between actual participation/familiarity and desire to participate/value. For example, a student who values traveling the world might not get a chance to do so at all during the semester; this does not necessarily change their computation of the value of travel. Behaviorally, actual HPW was not correlated with any measure, but desired HPW was negatively correlated with reaction time - as desired HPW increased, reaction time decreased. We interpret this result as indicating that subjects who place higher value on their preferred exemplar are more certain about their decision and therefore faster at making choices. Subjects choosing among less valued stimuli may tend to be less certain of their choices resulting in longer decision time.

Goal motivations are another factor that can be difficult to separate from value computation (O'Doherty, 2011, 2014). The value of a choice option like a gamble may be inherently tied to a goal such as maximizing payout. Here, subjects chose among category exemplars with no expectation of reward and no actual reward, such as a travel voucher (subjects were compensated with course credit for their participation in the study but not for their choices). Some studies have explicitly asked subjects to simulate reward experiences like receiving money, resulting in activation in mPFC similar to that observed when subjects were given a more concrete reward (Bray et al., 2010 & Miyapuram et al., 2012). In this task, subjects responded within 2.6 seconds on average. Given the time required to read the two answer options, such as

"road trip or international travel," and then respond via button box, it seems unlikely that participants were mentally simulating the experience of these outcomes. However, subjects were not specfically asked whether or not they did so and thus the possibility cannot be ruled out entirely.

Self referential processing could also plausibly co-occur in mPFC along with value computation, especially given that the choices participants made were tailored to their actual interests (Johnson et al., 2005; Brosch et al., 2012; Kim & Johnson, 2014). To account for this, participants were explicitly asked to report how much their top three stimuli related to their self concept, specifically whether they felt they would "not be the same person" if they could not play their favorite sport or engage in their favorite form of travel. Subjects average ratings of their #1 preferred stimulus were 2.9 out of 4, indicating the stimuli were "moderately related" to their identity. These ratings were not related to any brain activation above threshold, suggesting that self concept motivations were not driving mPFC coding. However,it is unclear if stimuli had been more closely customized to participant's preferences, whether these ratings would be higher and might be related to brain response in mPFC or other areas.

Salience is another factor that has been identified as a confound in many studies of value. Since items that are desired more are nearly always also more arousing or salient (such as larger potential monetary wins or more appealing foods), value and salience are often difficult to separate when using appetitive stimuli. However, salience can be defined several ways, such as probability/novelty (Ogawa et al., 2013), a motivational/attentional factor (Shiner et al., 2014), or absolute value (Litt et al., 2011; Kahnt et al., 2014). Using probability or absolute value as definitions of salience, our stimuli vary enormously in these factors. That is, a cruise vacation might be more salient than a road trip because of the novelty of travelling to a far away location, or greater financial cost leading to a higher absolute value of the trip. This type of salience may be entirely separate from the value of a cruise vacation for a subject who is afraid of water. That

said, we did not explicitly measure salience ratings and thus cannot completely rule out the possibility that salience may factor into some subject's value computations, particularly if salience is defined in as a motivational factor that guides attention. Future studies should consider using aversive stimuli to clarify this dynamic.

Taken together, these results demonstrate for the first time that the value of abstract reinforcer choices are tracked by mPFC according to individual preference and independent of expected or actual consequence. mPFC is apparently able to compute the "common neural currency" signal (Chib et al., 2009) even when the choice set is heterogeneous, abstract, and hypothetical.

Study 2: Brain response during choices for aversive abstract reinforcers

There are two possible mechanisms that could account for how the brain responds when making decisions about aversive stimuli. The salience account predicts increases in magnitude of voxel response related to stimulus intensity, such as large monetary wins or losses. The valence account instead predicts increases in voxel response for appetitive stimuli, and decreases for negative, a process akin to prediction error coding. Support in favor of both views has been found in studies using concrete aversive reinforcers. However, it remains unknown how factors such as choice outcomes, expectations, and goal motivations influence brain response during concrete aversive choices. In this study, participants choose among hypothetical aversive reinforcers such as types of illnesses or car accidents, which meet the same definition of "abstract reinforcer" used in Studies 1a and 1b. Using this approach, decisions for real world scenarios can be tested while avoiding potential confounds such as the hedonic response to painful stimuli like shocks, or the logistical challenges of implementing monetary losses in the lab. Additionally, we test how the framing effect, a bias well established in the concrete domain and recently in the appetitive abstract domain as well, affects behavior and brain response during aversive choice.

We make several predictions about the general effects of stimulus valence on choice: we expect that consistent with response in the appetitive domain in Study 1a, changes in activation will be observed in regions of the decision making network such as the striatum, mPFC, insula, and amygdala. However, we expect to observe primarily deactivation in these regions, in line with the valence account of brain response to aversive stimuli. Behaviorally, we expect that avoidance frames will result in faster decision times than approach frames, under the assumption that it is easier to decide which stimulus to avoid than approach. We also predict that differences by frame will be observed in patterns of brain interaction, following from results observed in the appetitive domain in Study 1a. Specifically, we expect to observe activation changes between the approach and avoidance conditions in limbic regions such as the striatum, insula, and amygdala.

Methods

Participants

Fourteen healthy adult participants (mean age= 24.43, SD= 4.9, 9 female) underwent functional MRI conducted at the Rutgers University Brain Imaging Center (RUBIC). Participants met standard MRI exclusion criteria (e.g.,. no metal implants, pregnancy, neurological disorders). Participants were recruited from the Rutgers University Newark community through a department based subject recruitment system and word of mouth. Undergraduates were awarded course credit for participation. One participant was left handed and was included in the analysis after a review of their data showed reaction time within the range of the sample and patterns of brain activation consistent with the rest of the group. All participants gave informed consent to participate. The study was approved by the Rutgers Institutional Review Board.

Procedure

A version of the Abstract Reinforcer Task (ART; Mills-Finnerty et al., 2014) with aversive categories was developed through behavioral piloting with an independent group of subjects (n=49) to determine an appropriate range of categories, exemplars within those categories, and to optimize task format. Participants selected from a set of four categories: illnesses, car accidents, train incidents, and house incidents. A full list of category examples is provided in the appendix. Participants were asked to select one category they found the most negative, and one they found least negative. Participants unsure of how to select the most negative category were given the additional instruction to select the category "they are most afraid of, or would least like to happen to them." Categories chosen as most negative by participants were car accidents (6), train incidents (5), and illnesses (3). No subjects chose house incidents. Each category contained 24 exemplars, 12 of which were deemed high severity and 12 low severity. High severity stimuli were defined as conditions that could lead to death (i.e. cancer, bomb threat on a train, house fire, head on car collision). Low severity stimuli were all

non-life threatening but still aversive (i.e. a paper cut, house vandalism, panhandlers on the train, paint scratch on a car).

In the scanner, participants made two-alternative forced choices between all possible combinations of category exemplars (i.e. "flu versus cancer"), once with the prompt "which would you rather avoid" (avoidance frame) and once as "which would you rather have" (approach frame). Participants completed two task runs, one where they chose amongst high severity stimuli and one where they chose amongst low severity. Each scan run took approximately 13 minutes. Choices were presented in 28 second blocks with 7 choices per block (except for the final block of each framing condition which contained 10 stimuli), and participants were given up to 4 seconds to respond. After participants selected their answer the screen changed to a crosshair to indicate the response had been logged. Twelve second rest periods divided the approach and avoidance blocks. As a behavioral control condition to account for baseline reaction time differences, participants made positive or negatively framed size judgments (e.g. "which is bigger/smaller, pebble or boulder?") over a 30 second continuous block. Stimuli were presented and responses recorded using PsychoPy (http://www.psychopy.org/). Participants also completed a similarity rating task which is described in the methods section of Study 3. Order of conditions was counterbalanced across participants, with 7 participants making choices about high severity stimuli first and low severity second and 7 choosing amongst low first and high second. Task order was also counterbalanced so that 8 subjects completed aversive framing, then similarity ratings, and 6 completed similarity prior to aversive framing.

Participants also completed the DOSPERT (Weber & Johnston, 2008) to measure risk behaviors and attitudes and familiarity ratings for all stimuli in their chosen category ranging from 1=never heard of it to 7=highly familiar.

Scanning Parameters

Functional imaging was conducted using a Siemens 3.0 Tesla Trio MRI scanner to acquire gradient echo T2*-weighted echo-planer (EPI) images with BOLD contrast. A 12 channel array coil was used due to increased signal detection in orbitofrontal regions. Each volume collected had 32 axial slices. 393 measurements were acquired in ascending contiguous order with a TR of 2s, for a total scan time of 13 minutes and 6 seconds. Imaging parameters included: field of view, 192 mm; slice thickness, 3mm; TR, 2s; TE, 30ms; flip angle, 90 degrees. Whole brain high resolution structural scans were acquired at 1 X 1 X 1 mm using an MP-RAGE pulse sequence.

fMRI General Linear Model

Analysis was performed using FMRIB's Software Library (www.fmrib.ox.ac.uk/fsl). Skull stripping was performed using BET (Brain Extraction Tool) and then individual data was registered to the anatomical standard using FLIRT (FSL's Linear Registration Tool), in which the BOLD functional data are registered to the MPRAGE anatomical scan and then to the MNI atlas image. FEAT (FSL's Expert Analysis Tool) was used for all GLM analysis with the following parameters for first level (individual scan) analysis: motion correction with MCFLIRT; 5 mm FWHM spatial smoothing, highpass filtering, and a second registration to the MNI atlas using 3 DOF. The two regressors used in first level analysis were the timepoints associated with the approach and avoidance frames; rest periods were used as baseline and therefore not modelled. The same processing steps were used for both the high severity and low severity approach and avoidance conditions at the first level.

At the second level, activation was modelled several ways: as the average above baseline magnitude (activation) and below baseline magnitude (deactivation) of each framing condition (approach and avoid); as a t test of the differences between activation in the approach and avoid conditions; and the average group activation with approach and avoidance collapsed together. This collapsing was done by modelling each subject's approach and avoidance related timepoints

together, producing individual files representing the average combined activation of the approach and avoidance conditions, referred to here as the "all high" condition. The same analysis procedures were used for the high and low severity conditions, to also create an "all low" condition. All group models were run using the Flame 1 mixed effects model and a cluster threshold of $z=1.65$, $p>.05$ unless otherwise stated.

High and low severity conditions were also compared using the average individual activation for high avoidance contrasted with low avoidance, high approach contrasted with low approach, and vice versa (low vs. high avoidance and approach). Average activation for both approach and avoidance collapsed together was also used to model the main effect of stimulus severity.

Finally, the following individual variables were used as regressors to predict brain activation in the high severity condition: familiarity ratings for stimuli, reaction time, risk attitude scores, and risk likelihood scores (the latter two taken from the DOSPERT; Weber & Johnston, 2008). All scores were demeaned prior to being entered as regressors. In order to clarify observed effects of reaction time on putamen activation during high severity framing (see results section), data were reanalyzed at the individual subject level using two predictors: activation during the first 64 trials (early trials) and activation during the last 68 trials (late trials), in order to model potential learning effects. The familiarity and RT regressors were then used to predict activation that was greater during the early compared to late period, as well as general activation effects of early>late and late> early.

Connectivity

Connectivity analysis was performed to quantify how brain network response during decisions for abstract aversive reinforcers is influenced by framing. While general linear model analysis addresses how conditions can affect the level of response by various brain regions, it can not reveal how those brain regions interact. Here, we use an Independent Multi-sample Greedy

Equivalence Search (IMaGES) (refer to Connectivity Methods in Study 1a). The algorithm starts with an empty graph and searches forward, one new connection at a time, until it finds the set of connections that optimally represents the entire group of subjects, interpolating any missing data. The algorithm searches with the restriction of finding only Markov equivalence classes of directed acyclic graphs, and without the option of varying time lags, given that doing so did not improve accuracy in simulation (see Ramsey et. al. 2010, 2011 for a more in depth methods discussion). The process is penalized to prevent overfitting using the Bayes Information Criterion (Schwarz, 1978): $-2\ln(ML) + k \ln(n)$, where ML is the maximum likelihood estimate, k is the dimension of the model (the number of directed edges plus the number of variables), and n is the sample size (number of participants). As the reliability of the IMaGES algorithm was found to be higher when using the LOFS search post filter, which determines and assigns the "dominant" direction of the edge (removing bidirectional edges), the current analysis was run using this option and thus only unidirectional edges were returned.(Ramsey et. al. 2011).

ROIs were chosen based on activation during GLM analysis. Binary masks were created for VMPFC and bilateral putamen using FSL view and the Harvard-Oxford anatomical atlas, in which the probabilistic atlas defined ROIs were converted into masks. Since activation both above and below baseline were observed using GLM analysis, regions where both activations and deactivations occurred were masked using more conservative methods. Specifically, the hippocampus mask was thresholded to 70% anatomical probability to exclude activation likely to be situated in other regions. For the insula, anterior cingulate, and amygdala, coordinates were restricted to those that fell within <70% probability of being a part of that region, and were then selected using the approximate center of the clusters active or deactive during GLM. A 9mm sphere was then created to mask that activation. Mask coordinates were chosen to ensure minimal overlap of active and deactive voxels (>5 voxels). For the insula, two masks were created to account for both activations and deactivation, one in anterior insula (activation) and one in poster

(deactivation). No voxel overlap occurred in the anterior cingulate, hippocampus, or amygdala, and minimal overlap (approx. 3 voxels) was observed for the insula and hippocampus masks. Actual masks used in analysis are pictured in Figure 4.1.

Average time series for each subject were extracted from these ROIs using FSL's meanTS module. The first and last TR of all condition blocks after the first block were excluded from analysis to exclude any carry over effects resulting from the hemodynamic response function time lag. Time courses of interest were arranged into a matrix for each subject, with the ROIs as columns and each row representing a single time point. These files were then input into the IMaGES workflow in Tetrad. IMaGES outputs a set of graphs that are all equivalently likely called a Markov Equivalence Class (MEC). Final graphs were selected by choosing the most complex graph for each condition (those with the most edges) within the MEC. Edge (connection) weights were exported from Tetrad into LibreOffice Calc (https://www.libreoffice.org/). T statistics were averaged across the group, and were used instead of raw coefficient values because they take into account standard error. The TDIST function was used to calculate significance values.

Results

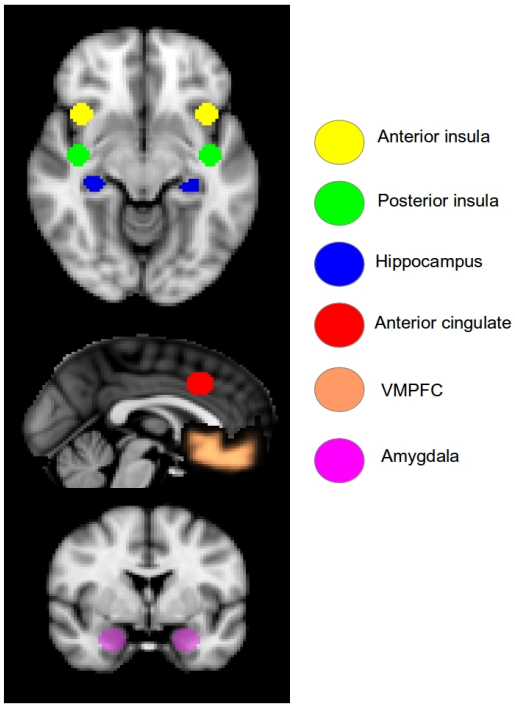


Figure 4.1. Masks used for timeseries extraction.

Behavioral

Reaction time was significantly longer for the approach (M=2.16, SD=.29) compared to the avoid condition for high severity stimuli (M=1.99, SD=.34; t = -6.3812, df = 13, p=.00002).

Reaction time was also faster for aversive framing compared to the appetitive framing reaction time reported in Study 1, for both positive/avoidance framing (F=9.12, p=.005) and negative/approach framing (F=13.55, p=.001).

General Linear Model

Conditions with results reported in detail here are listed in Table 4.1, pictured below.

Frame	Severity	Magnitude	Contrast	Active regions
Avoid	High	active	HighAvoid>HighApproach	right insula, right postcentral gyrus, and left caudate
Avoid	High	active	Average w/RT	bilateral putamen, bilateral hippocampus, left amygdala, brainstem, and right insula
Avoid	High	active	Average w/Risk likelihood	supramarginal gyrus, postcentral gyrus, middle temporal lobe, and occipital cortex
Avoid	High	active	Average w/familiarity	postcentral gyrus, superior parietal lobule, supramarginal gyrus
Approach	High	active	Average w/Risk likelihood	frontal pole, middle frontal gyrus, supramarginal gyrus, postcentral gyrus, middle temporal lobe, and occipital cortex
Approach	High	active	Average w/RT	bilateral putamen, bilateral hippocampus, brainstem, bilateral caudate, thalamus, middle temporal gyrus, temporal pole, and during avoidance framing in bilateral putamen, bilateral hippocampus, left amygdala, brainstem, and right insula
Approach	High	active	Average w/familiarity	postcentral gyrus, superior parietal lobule, supramarginal gyrus
All	High	active	average	right dorsal caudate, bilateral thalamus, pre- and post-central gyrus, supplementary motor area, anterior cingulate, lateral occipital cortex, superior parietal lobule, angular gyrus, middle temporal gyrus, and left hippocampus
All	High	active	average w/RT	left putamen

All	High	active	average w/ risk likelihood	supramarginal gyrus and postcentral gyrus
All	High	active	average w/familiarity	postcentral gyrus, superior parietal lobule, superior temporal gyrus, planum temporale
All	High	deactive	average	right insula, VMPFC, posterior cingulate, superior parietal lobule, right supramarginal gyrus, and right postcentral gyrus
Approach	Low	active	average	left putamen, left thalamus, left pre- and post-central gyrus, bilateral superior parietal lobule, bilateral lateral occipital cortex, occipital fusiform, anterior cingulate, paracingulate, bilateral middle frontal and inferior frontal gyrus, left insula, and left putamen
Avoid	Low	active	average	left thalamus, left pre- and postcentral gyrus, bilateral superior parietal lobule, bilateral lateral occipital cortex, occipital fusiform, anterior cingulate, paracingulate, and left middle frontal and inferior frontal gyrus
All	Low	active	average	left putamen, left thalamus, left insula, anterior cingulate/paracingulate, right frontal pole, bilateral middle and inferior frontal gyrus, left pre and postcentral gyrus, occipital fusiform gyrus
All	High/Low	active	HighAll>LowAll	insula, caudate, and putamen
All	High/Low	active	HighAvoid>LowAvoid	insula, inferior frontal gyrus, and superior temporal gyrus
N/A	High	active	Late>early average	precentral gyrus, SMA,

				postcentral gyrus, precuneus, anterior and posterior cingulate, ant. paracingulate, lingual gyrus, VMPFC, OFC, ;left caudate, bilateral insula, right amygdala
N/A	High	active	early>late w/RT	precentral gyrus, SMA, postcentral gyrus, precuneus, anterior and posterior cingulate, ant. paracingulate, lingual gyrus, VMPFC, OFC, ;left caudate, bilateral insula, right amygdala
N/A	High	active	early>late w/familiarity	precuneus, right LO, occipital pole, right cerebellum, supramarginal gyrus, bilateral postcentral, biL post cingm, inferior temporal gyrus

Significant activation was observed for the "all high" condition (collapsed across framing conditions), in the right dorsal caudate, bilateral thalamus, pre- and postcentral gyrus, supplementary motor area, anterior cingulate, lateral occipital cortex, superior parietal lobule, angular gyrus, middle temporal gyrus, and left hippocampus at a cluster threshold of $z=2.33$, $p<.05$ (Figure 4.2). No activation was greater for the direct comparison of the approach compared to avoidance frame. Greater activation was observed in the avoid>approach frame in the right insula, right postcentral gyrus, and left caudate using an Ordinary Least Squares regression with a cluster threshold of $z=2.33$, $p<.05$.

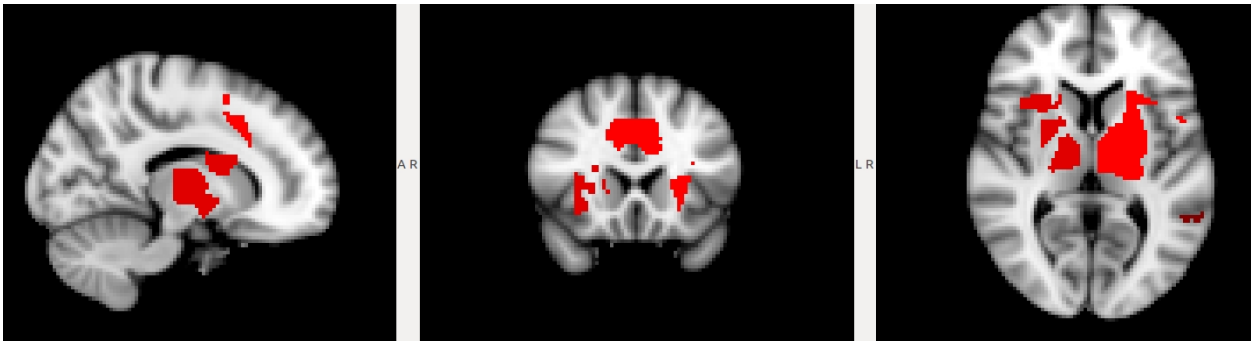


Figure 4.2. Activation for the all high condition.

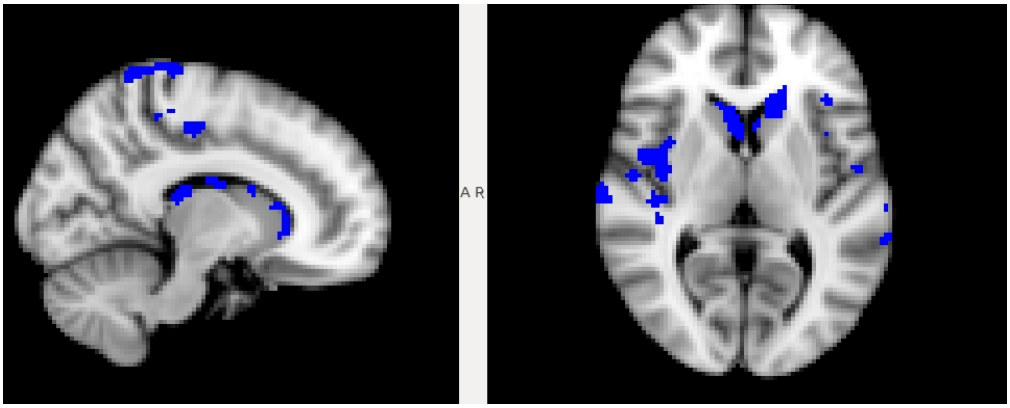


Figure 4.3. Activation for High Avoid> High Approach. Results shown at a cluster threshold of z= 1.25, p<.05 for visualization purposes.

Significant deactivations were also observed for the all high severity condition, in the right insula, VMPFC, posterior cingulate, superior parietal lobule, right supramarginal gyrus, and right postcentral gyrus at a cluster threshold of z=3, p<.05 (Figure 4.4). Deactive regions largely overlapped between the aversive and approach conditions, with the exception of clusters in right thalamus and subgenual posterior cingulate during the avoidance frame, and in superior temporal gyrus in the approach frame (z=2.33, p<.05). Differences in deactivation by frame were not compared directly due to conceptual and methodological impediments to comparing multiple deactive conditions. Instead, connectivity analysis was performed using both "active" and "deactive" nodes to clarify differences by frame type (see "connectivity results" section).

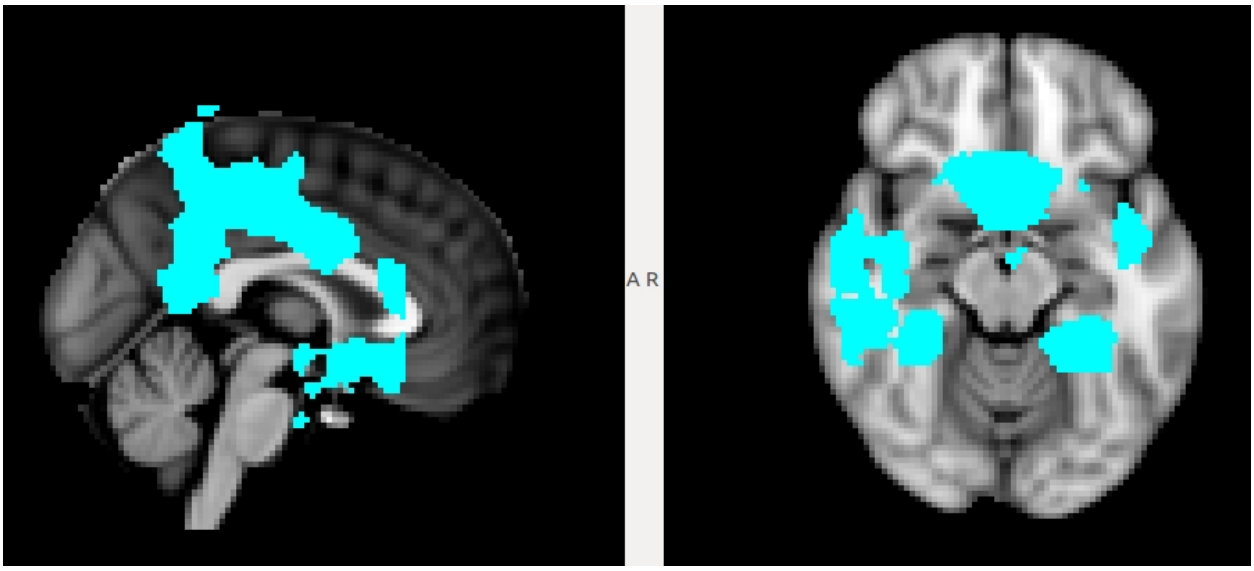


Figure 4.4. Deactivation during the all high condition.

Several scores derived from subject behavior and self-report questionnaires were also used as predictors of activation during high severity framing: global risk attitude and risk behavior likelihood scores, scores measuring degree of familiarity with aversive stimuli, and reaction time. Risk attitude scores did not predict any activation at a cluster threshold of $z=1.65$, $p<.05$ for high severity framing, either for approach and avoidance measured separately or for all high. Risk likelihood scores predicted activation during avoidance framing in supramarginal gyrus, postcentral gyrus, middle temporal lobe, and occipital cortex, and during approach framing in frontal pole, middle frontal gyrus, supramarginal gyrus, postcentral gyrus, middle temporal lobe, and occipital cortex. For all high, risk likelihood scores predicted activation in supramarginal gyrus and post-central gyrus. Familiarity scores for aversive stimuli predicted activation during both the approach and avoidance conditions in postcentral gyrus, superior parietal lobule, and supramarginal gyrus at a cluster threshold of $z=2.33$, $p<.05$. For the high all condition with familiarity as a predictor, activation was observed in postcentral gyrus, superior parietal lobule, superior temporal gyrus, planum temporale at a cluster threshold of $z=2.33$, $p<.05$.

Reaction time scores predicted activation during approach framing in bilateral putamen, bilateral hippocampus, brainstem, bilateral caudate, thalamus, middle temporal gyrus, temporal

pole, and during avoidance framing in bilateral putamen, bilateral hippocampus, left amygdala, brainstem, and right insula. Activation predicted by RT for all high was observed in the left putamen at cluster threshold of z=2.33, p<.05 (Figure 4.5). When the effects of RT were compared directly for approach>avoidance and avoidance>approach framing, no significant effects were observed at a cluster threshold of z=1.65, p<.05.

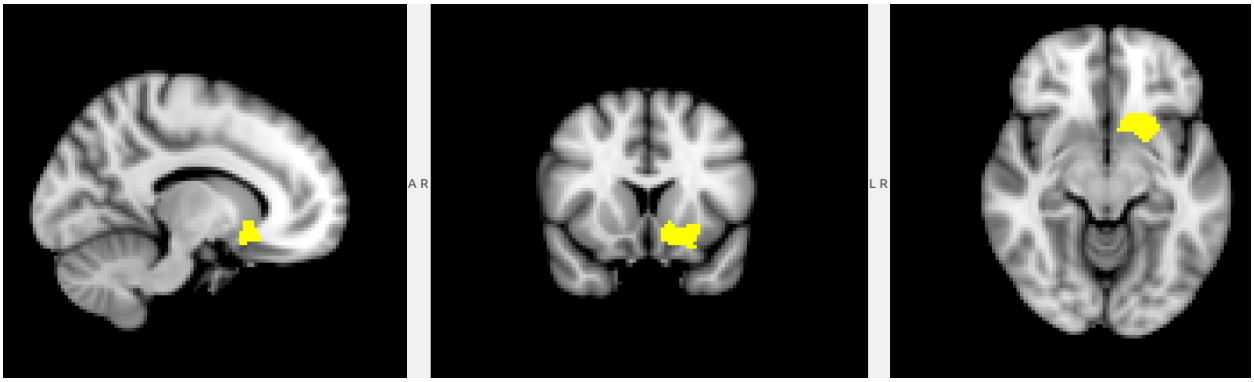


Figure 4.5. Activation predicted by reaction time in the all high condition.

In order to clarify the effect of reaction time on putamen activation during high severity framing, trials were split into early and late in order to study potential learning effects (refer to methods section). Scores for familiarity as well as reaction time were used as predictors of activation during the contrast of early>late and late>early. When early and late were compared without any other predictors in the model, greater activation was observed for the late>early contrast only in precentral gyrus, supplementary motor area, postcentral gyrus, precuneus, anterior and posterior cingulate, anterior paracingulate, lingual gyrus, VMPFC, OFC,left caudate, bilateral insula, and right amygdala. During early>late with familiarity as a predictor, activation was seen in precuneus, right lateral occipital, occipital pole, right cerebellum, supramarginal gyrus, bilateral postcentral, bilateral posterior cingulate, and inferior temporal gyrus. During early>late with RT as a predictor, activity was observed in left putamen, left accumbens, left OFC, left IFG, left angular gyrus, occipital fusiform, bilateral cerebellum,and left frontal pole. At a cluster threshold of z=2.33, activation predicted by RT during early>late was observed in the

same region as when RT was used to predict activation during All High. An overlay of the RT effect during All High and early>late is shown in Figure 4.6. No activation was seen for the late>early contrast with either familiarity or RT as a predictor.

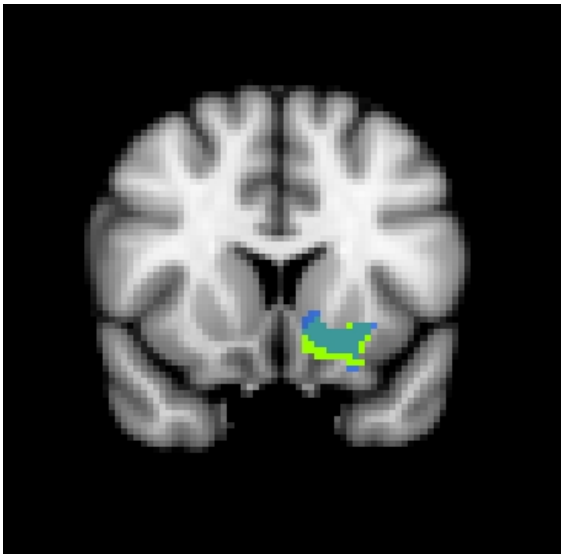


Figure 4.6. Dark blue=effect of RT during early>late; green = effect of RT during All high; teal = overlapping voxels.

During the "all low" average for low severity choices, activation was observed in left putamen, left thalamus, left insula, anterior cingulate and paracingulate, right frontal pole, bilateral middle and inferior frontal gyrus, left pre- and postcentral gyrus, and occipital fusiform gyrus at a cluster threshold of z=2.33, p>.05 (Figure 4.7). During the group average of the low severity approach condition, activation was observed in left putamen, left thalamus, left pre- and post-central gyrus, bilateral superior parietal lobule, bilateral lateral occipital cortex, occipital fusiform, anterior cingulate, paracingulate, bilateral middle frontal and inferior frontal gyrus, left insula, and left putamen . During the group average of the low severity avoid condition, activation was observed in left thalamus, left pre- and post-central gyrus, bilateral superior parietal lobule, bilateral lateral occipital cortex, occipital fusiform, anterior cingulate, paracingulate, and left middle frontal and inferior frontal gyrus. During the direct t-test of

approach>avoid and avoid>approach for low severity stimuli, no activation was observed above a cluster threshold of z=1.65, p<.05.

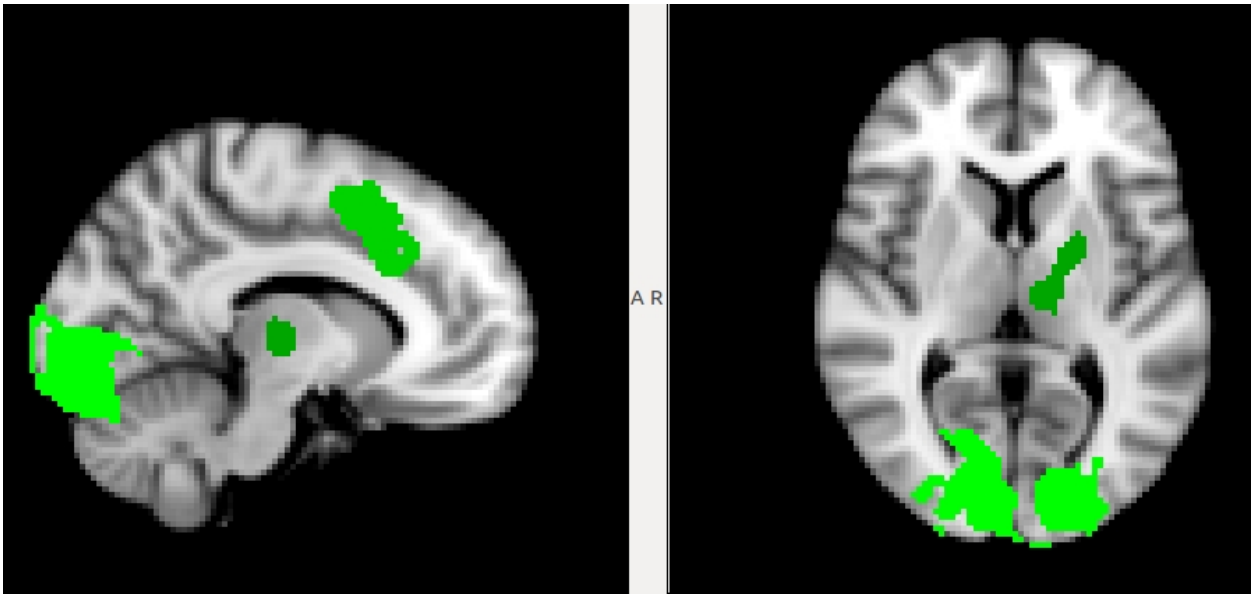


Figure 4.7. Average activation during the all low severity condition.

Activation during the high severity condition was also compared to that during the low severity condition. For the contrast of high avoidance framing >low avoidance framing, activation was seen in the insula, inferior frontal gyrus, and superior temporal gyrus at a cluster threshold of z=2.33, p<.05 (Figure 4.8). For the contrast of high approach framing >low approach framing, no activation was observed above a cluster threshold of z=1.65, p<.05. For the contrast of highall>low all, activation was observed in the insula, caudate, and putamen. For all contrasts of the low severity condition>high severity condition, no activation was observed above a cluster threshold of z=1.65, p<.05 (Figure 4.8).

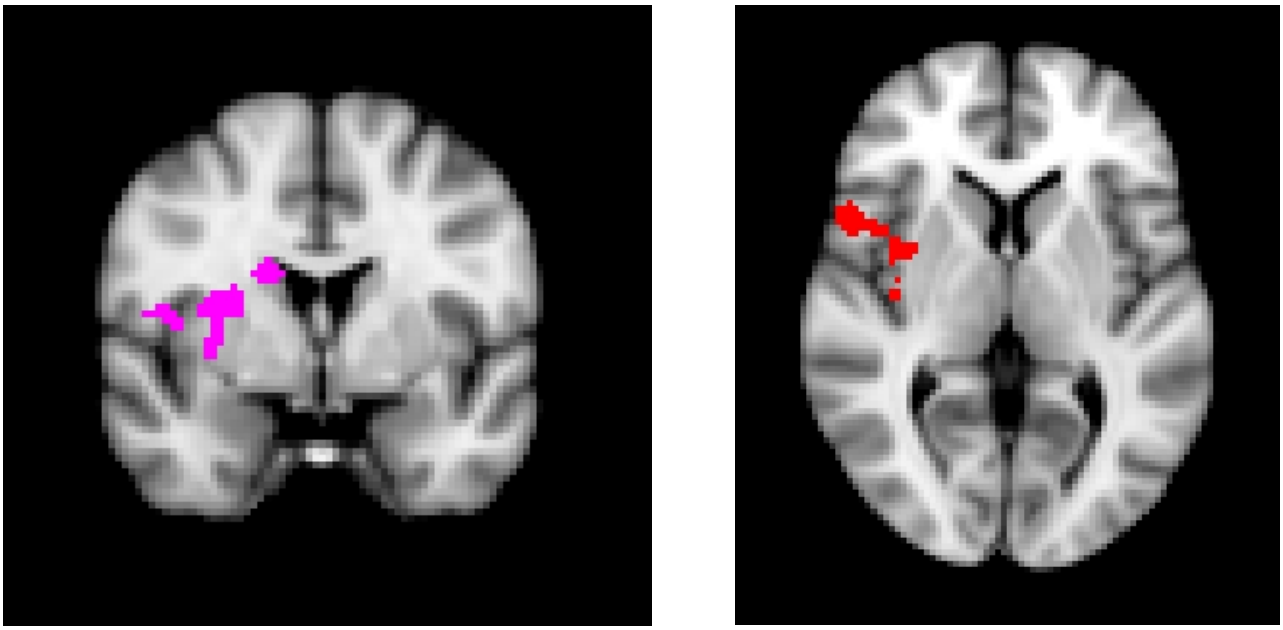
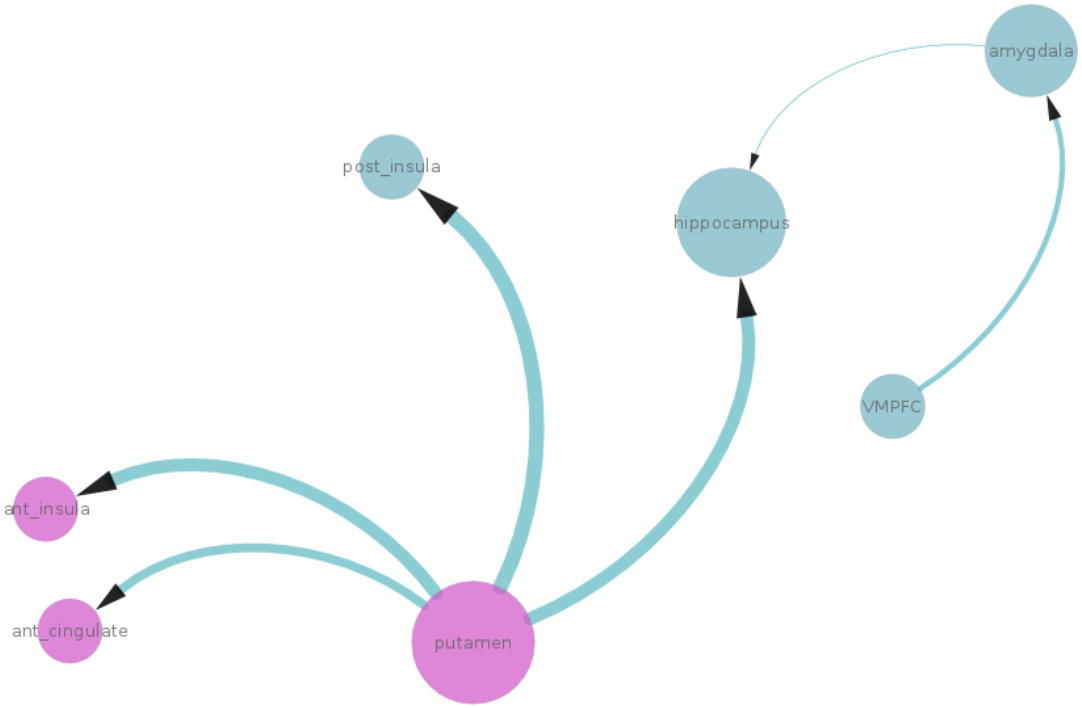


Figure 4.8. Activation for the contrast of High All > Low All (left, in pink) and HighAvoid > Low Avoid (right, red).

Results, Connectivity

Approach and avoidance related connectivity was measured separately in the following network of regions: putamen, anterior insula, and anterior cingulate (areas active above baseline); and posterior insula, VMPFC, hippocampus, and amygdala (areas active below baseline). A connection to B, originating from A, is indicated here as A->B, whereas a connection from B to A is indicated as B->A. During both avoidance and approach framing, the following connections were observed: putamen->anterior insula, putamen->anterior cingulate, putamen->posterior insula, putamen->hippocampus, VMPFC->amygdala, amygdala->hippocampus. During avoidance framing, additional connections were observed from posterior insula->amygdala and putamen->VMPFC. All connections were significant at $p<.005$, except for the posterior insula->amygdala connection ($p=.07$) and putamen->VMPFC ($p<.03$) in avoidance framing. Connection strength (coefficients) did not differ significantly for the connections common to both approach

and avoid framing. Complexity as measured by beta index was higher for avoidance framing (beta=1.14) than approach framing (beta=.86).



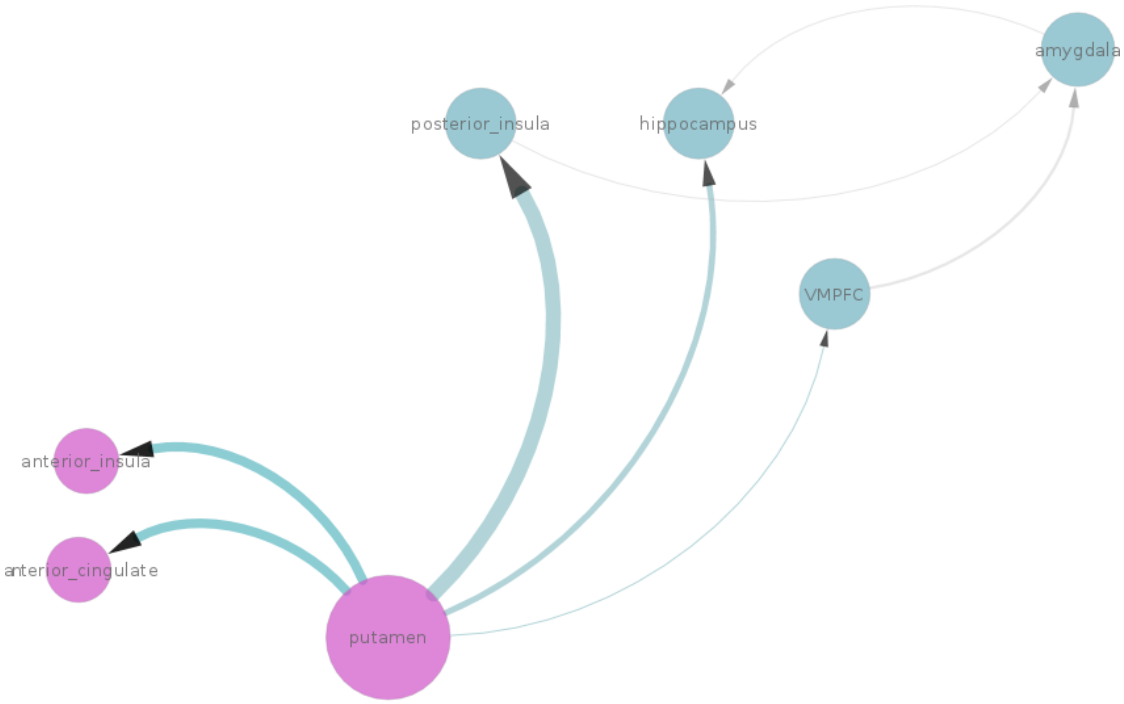


Figure 4.9. Connectivity analysis of approach and avoidance framing. Connectivity changes related to frame included increased connection amongst deactive regions (amydala and posterior insula) and from active to deactive regions (putamen to VMPFC). Node color: pink=active region, blue=deactive region. Edge thickness: thicker=higher coefficient. Arrows indicate direction of connection.

Discussion

The present study aims to characterize brain response to aversive abstract reinforcers. The manipulations used include framing of choice prompts (approach vs. avoidance), severity of aversive stimuli (high vs. low), and three additional predictors of activation - reaction time, familiarity ratings, and risk behavior scores. Overall, the results support a valence coding account of brain response to aversive abstract reinforcers. In the high severity condition, more deactivation was observed than activation within regions associated with decision making (in terms of magnitude and voxel extent of response), suggesting that the primary mechanism of

response to aversive abstract reinforcers is deactivation. The hypothesis that choosing which aversive stimulus to avoid would be processed similarly to choosing which appetitive stimulus to approach was supported, via greater activation and network connectivity for avoidance>approach framing within reward sensitive regions. Additionally, results related to reaction time effects, familiarity, and risk attitudes suggest that each of these factors uniquely contributes to decisions for aversive abstract reinforcers.

Framing and aversive abstract reinforcer response magnitude

Framing effects have been robustly observed in the concrete context (e.g. Kahneman & Tversky 1986; refer to Kuhberger 1998 for meta-analysis), and recently established in the abstract context as well (Mills-Finnerty et. al., 2014). However, no studies to our knowledge have tested the effects of approach and avoidance frames on choices for hypothetical aversive abstract reinforcers. Consistent with predictions and the existing literature (e.g. C. Alos-Ferrer et. al., 2012, Foo et. al., 2014), significant reaction time differences were observed for approach versus avoidance frames in both the high and low severity conditions, with significantly faster RT for avoidance compared to approach. Since RT is typically interpreted as an index of task difficulty, these results suggest that choosing which aversive AR to approach is more difficult than choosing which one to avoid. Since avoidance is the more positive or desirable outcome, it follows that these choices can be made more quickly and easily. Reaction time was also faster for all conditions of the aversive framing task compared to the behavior during appetitive framing reported in Study 1a. We interpret the differences between reaction time in aversive vs. appetitive framing as indicating that aversive stimuli may induce a "fight or flight" type response that motivates faster responding; that is, fleeing from a bear has a stronger time imperative than deciding which hiking path to take.

Despite highly significant behavioral effects, no differences in brain activation were observed for the direct contrast of approach>avoidance for either high or low severity stimuli.

The results of avoidance>approach were evident in the high severity condition using an Ordinary Least Squares regression. This method is less rigorous than the FEAT Flame 1 mixed effects model used for all other analyses, but is still widely used in the field (Mumford & Nichols, 2009), and was implemented in conjunction with a more stringent cluster threshold of $z=2.33$, $p<.05$ to reduce risk of false positives. Average activation for approach and avoidance, for both high and low severity stimuli, largely occurred in overlapping regions. Thus, it appears that the framing manipulation has smaller effects on magnitude increases in the context of aversive choices. There are several possibilities why this may be the case: differences in task design (i.e. event related versus block) or insufficient sample size may have contributed to the lack of significant differences. If changes in magnitude in response to framing effects in the aversive domain are more subtle than in the appetitive domain, a sample of 20 or 25 participants might be required to detect differences. Another significant issue involves the presence of deactivations during the task. Since comparing multiple deactive conditions poses methodological challenges not addressed well by current software, it remains unclear if there are frame based differences in deactivation magnitude during choices for high severity stimuli. However, when deactivations were measured in terms of network connectivity, frame based differences were observed using connectivity analysis (see next section).

Another explanation for the reduced effect of framing on brain response involves the interaction between framing effects and anchoring effects (e.g. Wu & Cheng, 2011; see Furnham 2011 for review). Anchoring effects refer to the tendency of subjects to "anchor" their answers to irrelevant values during decision making. In appetitive framing, even the negative frame ("which do you like less"), which may be perceived as rejecting one of the two options, implies preference for the unselected option. Thus the anchor or reference point for decisions is either positive (select the one you like best, even if you like both) or neutral to positive (reject the one you like less, but "keep" the one you like more). For example for "hiking vs. biking" rejecting biking

implies a preference for hiking. For aversive stimuli, on the other hand, for "paper cut vs. brain freeze," even if you reject the paper cut you are still stuck with brain freeze. There is no positive or neutral interpretation of any choices made in either the approach or avoidance frames. It is possible that at the level of brain activation, variance related to perceived neutral or positive choices is a necessary comparison point to differentiate response to framing in the aversive domain. Clarifying if this is the case would require a framing task that presents positive, neutral, and aversive choices both as separate conditions (positive stimuli only, neutral only, and negative only, each framed as positive or negative decisions) and direct comparisons across valence (i.e. positive vs. neutral, negative vs. neutral, negative vs. positive).

Connectivity dynamics underlying framing effects in the aversive domain

In contrast to the GLM results, effects of approach vs. avoidance frame were observed via connectivity analysis and shed light on differences between processing of appetitive and aversive abstract reinforcers. Many of the areas that showed greater activation during appetitive framing in Study 1a exhibited significant deactivation during aversive framing, including the insula and mPFC. Differences in deactivation patterns could not be measured directly using currently available GLM methods; however, results from connectivity analysis suggest frame-based differences in deactivation.

The putamen appears to play a central role in both activation and deactivation networks during both the approach and avoidance frames. The putamen had the most connections of any region in the network during both conditions. There were more connections between the putamen and several deactive regions (posterior insula and VMPFC) during avoidance, but not approach framing. GLM analysis of reaction time effects, familiarity ratings, and putamen activity suggest that some of its role may involve learning dynamics (see subsequent section, "Learning effects during aversive abstract reinforcer choices"). However, this dynamic is more prominent during

early trials compared to trials later in the task, and thus is unlikely to be the only role of the putamen during the task.

Results from the literature suggest that aversive prediction error responses are coded by regions of caudate and putamen (e.g. Gottfreid et al., 2002: O'Doherty et al., 2006; Delgado et al., 2008; see Bissonnette et al., 2014 for review). Several studies have used both appetitive and aversive stimuli to measure PE. For example, one study found that the putamen, in addition to the anterior insula and rostral anterior cingulate, was responsive during prediction errors involving both unexpected relief and exacerbation of pain (Seymour et al., 2005). Interestingly, the specific sub-regions of the striatum, insula, and anterior cingulate that decreased activation in response to prediction error in Seymour et al. (2005) were active during choices for high severity stimuli in our study, whereas the posterior insula and posterior cingulate both contained deactive voxels. In another study that used high resolution imaging (Mattfield et. al., 2011), the region of caudate that is active for positive PE (right caudate head) is deactive during the all high aversive condition in our results. The more anterior portion of the caudate that showed greater deactivation during negative PE in their study had greater activation in ours. These results suggest that the same regions that are involved more generally in PE are active or deactive during our task. However, without high resolution imaging and given the differences in protocols, it is difficult to interpret how meaningful small differences in voxel cluster location are. Further, since there are no expectations or outcomes in our task, it is unlikely that putamen activation or connectivity represents prediction error. However it is possible that the putamen codes the hypothetical negative outcomes associated with choices. The consistent pattern of magnitude increases within regions of putamen, anterior insula, and caudate for aversive stimuli suggest that in this context, some sub-regions of these areas may be more sensitive to intensity or salience rather than value. Value may instead be coded in deactive regions, such as other sub-regions of putamen, or in VMPFC and/or posterior insula, which receive potentially modulatory inputs from the putamen.

To better clarify value versus salience dynamics, in future studies participants could explicitly rate each of these factors, ideally after every choice. However, the primarily deactivation-based dynamics in Study 2, especially in contrast to the activation centric results of Study 1, provide strong support for deactivation based, valence sensitive processes during choices for aversive abstract reinforcers (see "Conclusions" section for further discussion).

The sub-network of deactive regions ("deactivation network") had more intra-connections amongst regions than the activation network in both framing conditions. This deactivation network connectivity increased substantially during avoidance framing, with two unique connections (putamen->VMPFC, posterior insula->amygdala). This increase in deactive network connectivity in limbic regions for avoidance compared to approach is in line with predictions regarding the brain response to avoiding a negative stimulus. Specifically, it was predicted that areas such as the putamen and mPFC which increase activation during choices for appetitive abstract reinforcers should decrease activation for aversive abstract reinforcers. This prediction was partially confirmed, in that putamen increased its activation for avoidance>approach frames, but mPFC decreased its activation. Connectivity results suggest that the decreases in mPFC during avoidance framing may actually be driven directly by the increases in putamen activation. The putamen may code factors such as the hypothetical aversiveness of the choice options, or the intensity/perceived severity of those choices, information that may be incorporated into a value signal in mPFC.

The presence of activation and deactivation within different sub-regions of the same brain areas also suggests that potentially opponent processes are co-occuring in response to aversive stimuli. This delineation may be based on functional specializations of these subregions. In the high severity condition, activation was observed in the anterior insula and deactivation in the posterior insula. These sub-regions have been implicated in different aspects of interoception - anterior insula with cognitive and affective components (such as feelings of disgust) and posterior

insula with sensory encoding (such as the experience of pain; see review by Uddin, 2014). Interestingly, connectivity analysis revealed connections between the putamen and both anterior and posterior insula during both approach and avoiding high severity framing. During avoidance framing only, an additional connection from posterior insula->amgydala was also present. These results suggest that posterior insula is the sub-region more affected by the difference between approach and avoidance prompts for aversive stimuli. Further, the unique connection from posterior insula->amgydala during avoidance suggests that insula activation may serve to modulate or downregulate activation in the amygdala. Given the role of the amygdala in responding to aversive stimuli (e.g. O'Doherty, 2001; Whalen et al., 2004; Orsini et al., 2015), particularly during fear learning (e.g. Nader et al., 2000; Wolff et al., 2014; Moscarello et al., 2014) these results suggest that inputs from the posterior insula may directly influence this response, such as by relaying information about relevant sensory features of stimuli.

Interestingly, the connection weights for the posterior insula->amgydala connection were positive for half the subjects and negative for the other half. A negative coefficient indicates that as the posterior insula increases activation the amygdala decreases activation. The mix of positive and negative coefficients for this connection suggest several possibilities. One is that because IMaGES provides a summary snapshot of dynamics that may have different patterns at different points in time, that the connection from posterior insula->amgydala may be inhibitory at some points and excitatory at others. These processes may sum differently for different subjects. Or, the subjects with positive coefficients may actually have a different communication dynamic between these regions than the subjects with negative coefficients. It is possible with a sample closer to 25-30 that subgroups of inhibitors and exciters would emerge. However, an important caveat is that since degree of deactivation is considered a marker for value processes within analyses that measure magnitude, a negative graph coefficient does not indicate whether the dynamic represented is "less activation" (e.g., a reduction of involvement or communication) or "more

deactivation" (which could indicate an increase in involvement or communication). Resolving such questions would require tools to measure and compare magnitude and connectivity changes in deactivation, ideally over time.

Learning effects during aversive abstract reinforcer choices

An additional effect was observed when reaction time was used as a predictor of activation. In Study 1a, RT did not predict any activation or account for any task related variance. However during aversive framing, RT predicted activation during both the approach and avoidance conditions in a variety of regions including putamen, hippocampus, amygdala and insula, as well as in the all high condition in the putamen only. In order to test the hypothesis that this relationship between RT and putamen activation represents learning processes, the high severity trials were split into early (first half of the task) and late (second half). This approach presumes that activation that only occurs early in the task is likely to represent learning processes (e.g. Niznikiewicz & Delgado, 2011). When the early and late periods were compared directly, activation was only observed for late>early, suggesting that the overall task variance can not be explained as a learning or habituation effect. However, when RT and familiarity were used as predictors in the comparison of the early and late periods, significant activation only occured in the contrast of early>late. For early>late with RT, activation was observed in the same left putamen voxels as during the all high average with RT as a predictor. We interpret these results as indicating that reaction time and familiarity scores capture variance associated with learning during the beginning of the task. That is, some subjects reported being completely unfamiliar with a small number of stimuli (i.e. diseases such as Huntington's). When forced to choose between "Huntington's or cancer," these subjects must resolve the ambiguity of not knowing what Huntington's is is in some way, such as by creating a decision strategy, even an arbitrary one equivalent to flipping a coin. This may increase task difficulty, reflected by reaction time variance, until the strategy is established ("never pick Huntington's"). This account may explain

why familiarity and RT predict activation during early>late but not late>early. We intend to follow up these results in future analyses using probabilistic modelling of choice behavior, in order to quantify if and how choices become faster and more stable as subjects learn and enforce decision strategies.

Effects of stimulus severity of aversive abstract reinforcers

In line with predictions, the contrasts of all high vs. all low severity stimuli showed greater activation for high>low and no significant activation greater during low>high. Activation was observed in the insula, caudate, and putamen, areas that have been heavily implicated in aversive prediction error. Thus, this activation may represent greater anticipated hypothetical aversiveness of stimuli based on severity. Things that can kill you are worse that things that can't, and it is not surprising that this distinction is represented in the brain. What is notable is that while greater activation was observed in the insula, inferior frontal gyrus, and superior temporal gyrus during HighAvoid>LowAvoid, no differences above threshold were observed for HighApproach>LowApproach. It is of course possible that there are differences not detected in the current sample size. However, the greater activation differences for avoidance but not approach suggest that the severity distinctions may be more meaningful, and therefore require more brain activation to process, in the avoidance domain. That is, forced to approach two bad options, the overwhelming aversiveness of the choice may lead to similar activation magnitude whether the choices are cancer vs. Parkinson's, or a headache versus a cold. Whereas, given the chance to escape one option, severity becomes more decision relevant because of the potential goal motivation to avoid whatever is the worst or most severe option.

Risk attitudes and aversive abstract reinforcer choices

Both risk attitudes and ratings of engagement in risky behaviors (such as skydiving, cheating on taxes, or engaging in unprotected sex) taken from the DOSPERT were used as predictors of activation during the high severity condition, but only risk behavior scores predicted activation. For the approach condition, this activation involved more prefrontal regions such as frontal pole and middle frontal gyrus, whereas during avoidance activation was observed in middle temporal gyrus and precentral gyrus. The activation of regions associated with cognitive control and executive function during the approach condition suggests that participant's real world likelihood to take risks affects the areas recruited when forced to approach an aversive stimulus. However, it is not clear if this effect is driven by the participants with higher or lower risk ratings. That is, participants less likely to engage in risky behavior might have greater frontal activation during approach choices, perhaps reflecting a difficulty effect. Participants who engage in more risky behaviors might have less activation because they perceive choosing something aversive to be less negative or difficult. Put another way, someone who already engages in risky health behaviors might be less sensitive to the "risk" of choosing cancer over HIV compared to a more cautious person. Resolving this question would require that participants be grouped into high and low risk taking groups, such as via a median split on risk behavior scores, an approach that requires a larger sample size. However, these results suggest that there is a relationship between real-world risk taking behavior and perception of aversive abstract reinforcer choices.

Conclusion

In sum, Study 2 provides evidence in support of a valence rather than salience coding account of brain dynamics during choices for aversive abstract reinforcers. Differences in response to approach and avoidance frames was detected with the most sensitivity using

connectivity analysis, which revealed increases in connectivity between active and deactive regions during avoidance framing compared to approach framing. Putamen activation was central to both network communication as well as magnitude based analyses, and some of its role may be supporting processes akin to learning decision rules. Severity of stimuli modulated activation in areas associated with aversive prediction error, despite the lack of action-outcome contingencies in the task, suggesting a broader role for these regions in coding aversive stimuli. Finally, risk behavior scores were related to activation during the approach condition, suggesting that real world behavior directly influences brain coding of hypothetical aversive choices.

Study 3: Abstract Reinforcer Similarity

When participants make choices for abstract reinforcers in the absence of action-outcome contingencies, there are any number of dimensions that can be used to evaluate and choose among stimuli. Unlike a concrete reinforcer (CR) task such as gambling where subjects base choices on probability and magnitude, abstract reinforcer (AR) tasks rely on previously learned complex preference structures for stimuli. Therefore, subjects have a number of dimensions that can be used, idiosyncratically, to make choices. For example, a vacation choice can involve various dimensions such as climate, culture, food, ease of access, language, etc. Despite the range of dimensions that can be used, and the individualized weighting of those dimensions, , results from Study 1a, 1b, and 2 suggest that AR related choices reliably activate a stable network of regions associated with decision making. In Study 3, we measure whether two particular dimensions - value and familiarity - influence decisions in AR tasks even when those dimensions are irrelevant to the task. We measure this using the framework of value based attentional capture (VBAC), a phenomena where choice-irrelevant value features of stimuli draw both attention and brain resources during a task (e.g. Fukuda & Vogel, 2009; Anderson et al., 2011a+b). We test two accounts of this dynamic: if value is calculated implicitly through bottom up processes during similarity judgements we expect subjective value ratings to predict activation during the task. If value of ARs only influences choices in explicitly value-relevant choice contexts, then subjective value ratings should not predict activation during similarity rating, a result that would instead support a top-down effect of VBAC for abstract reinforcers.

We also test several more general questions about the structure of abstract reinforcer choices. First we model activation related to the overall effect of judging similarity of ARs, as well as how brain response differs during high and low similarity judgements. We predict that similarity ratings should primarily predict activation in areas associated with categorical knowledge and visual processing, such as lateral occipital cortex and cuneus. We also investigate

whether activation patterns observed during preference judgements in Study 1a also appear during similarity ratings, in areas such as the hippocampus and striatum. Additionally, we measure whether familiarity with stimuli and subjective value influence brain activation during similarity judgements. We expect that both value and familiarity will predict activation during similarity ratings in areas implicated in such processes, such as striatum and mPFC, in line with a bottom up account of VBAC.

Methods

Participants

Twelve healthy adult participants (mean age= 23.75, SD= 5.3, 7 female) underwent functional MRI conducted at the Rutgers University Brain Imaging Center (RUBIC). Participants met standard MRI exclusion criteria (e.g.,. no metal implants, pregnancy, neurological disorders). Participants were recruited from the Rutgers University Newark community through a department based subject recruitment system and word of mouth. Undergraduates were awarded course credit for participation. One participant was left handed and was included in the analysis after a review of their data showed reaction time within the range of the sample and patterns of brain activation consistent with the rest of the group. All participants gave informed consent to participate. The study was approved by the Rutgers Institutional Review Board.

Procedure

Participants selected a category that represented something they liked and were familiar with from the following options: sports, cultural activities, active lifestyle, historical periods, architectural styles, travel, countries, and reading material. The categories used were the same as in Studies 1a and 1b. Categories chosen as preferred by subjects in this study included reading material (3 subjects), travel (3 subjects), countries (3), active lifestyle (2), and cultural activities (1).

In the scanner, participants rated the similarity of all possible combinations of category exemplars on a scale ranging from 1=not similar at all to 7=highly similar. During each trial, the prompt "Rate the similarity" and two answer choices appeared on the screen above a rating scale. Participants had up to 8 seconds to respond via button box before the next set of answer choices appeared. Once the participant responded, the screen changed to a crosshair fixation for the remainder of the trial. This individually determined fixation period served as an inter-trial interval of variable length to help maximize statistical independence of trials. Stimuli were presented via an event related design paradigm. For a behavioral control condition to measure reaction time differences, participants rated similarity of colors during two 32 second blocks. Stimuli were presented and responses recorded using PsychoPy (http://www.psychopy.org/).

After scanning, participants filled out a questionnaire in which they first ranked their preferences for the stimuli in their chosen category in order from 1-12, 1 being most preferred. Then, they answered the following questions about their involvement with their top three preferred high value category exemplars: "how often do you engage in this activity every week (i.e. reading this material, traveling in this manner, playing or watching this sport, studying this style of architecture or historical period)" to measure actual hours per week (aHPW) and "if time or money weren't an obstacle, how much would you like to engage in this activity each week?" for desired HPW (dHPW). Finally, in order to measure the relationship between participant's self concept and preferred stimuli, they were also asked "How much do you feel like engaging in this choice activity is part of your identity (i.e. you would not be the "same person" if you were unable to read your favorite reading material, do yoga or your favorite activity, play or watch your favorite sport, engage in your favorite form of travel, etc.)?" Participants provided a rating from 1-4 as follows: 1=Not a part of my identity at all 2=slightly part of my identity 3= moderately part of my identity 4=significant part of my identity.

Scanning Parameters

Functional imaging was conducted using a Siemens 3.0 Tesla Trio MRI scanner to acquire gradient echo T2*-weighted echo-planer (EPI) images with BOLD contrast. A 12 channel array coil was used due to increased signal detection in orbitofrontal regions. Each volume collected had 32 axial slices. 393 measurements were acquired in ascending contiguous order with a TR of 2s, for a total scan time of 13 minutes and 6 seconds. Imaging parameters included: field of view, 192 mm; slice thickness, 3mm; TR, 2s; TE, 30ms; flip angle, 90 degrees. Whole brain high resolution structural scans were acquired at 1 X 1 X 1 mm using an MP-RAGE pulse sequence.

fMRI General Linear Model

Analysis was performed using FMRIB's Software Library (www.fmrib.ox.ac.uk/fsl). Skull stripping was performed using BET (Brain Extraction Tool) and then individual data was registered to the anatomical standard using FLIRT (FSL's Linear Registration Tool), in which the BOLD functional data are registered to the MPRAGE anatomical scan and then to the MNI atlas image. FEAT (FSL's Expert Analysis Tool) was used for all GLM analysis with the following parameters for first level (individual scan) analysis: motion correction with MCFLIRT; 5 mm FWHM spatial smoothing, highpass filtering, and a second registration to the MNI atlas using 3 DOF.

For each participant, similarity ratings for all trials were entered as a trial-by-trial predictor of brain activation, referred to as the "all similarity" condition. Rest periods and color similarity ratings were used as baseline and therefore not modelled. An additional first level analysis was conducted in which each subject's trials were split into high and low ratings. For 9 participants, 1-3 were considered low ratings and 4-7 high. Overall, subjects rated more stimulus pairs as low similarity (M=24.4, SD=8.6) than high similarity (M=39.8, SD=9.74; t = -2.946, df = 11, p-value = 0.01). One subject rated the majority of stimuli as low severity and so their data was recoded as 1-2=low and 3-7=high, resulting in 42 low cases and 23 high cases. 2 subjects

included a very small number of "4" responses (6 for one subject, 2 for the other) and so their data was coded using the rule 1-4=low and 5-7=high. After trials were split into high and low, the time points associated with each subject's individual ratings were used as predictors of high and low similarity related brain activation.

At the second level, the group average activation for the all similarity condition was measured. Activation during high similarity ratings was also compared directly against that for low similarity ratings. Several covariates were also used as predictors of activation during all similarity: actual and desired hours per week (HPW) scores for the #1 preferred stimuli for all participants, and ratings of the relationship between this stimulus and participant self image (referred to as identity ratings). For the contrast of high and low similarity, RT was also used as a predictor.

Results

Behavioral

Reaction time did not significantly differ between high similarity ratings and low similarity ratings (mean high RT = 3.83, SD=.74, mean low RT =3.57, SD =.51; t = 1.7353, df = 11, p-value = 0.11). Reaction time differences were observed in an independent behavioral pilot sample (mean high RT = 3.04, SD=.73, mean low RT =2.7, SD =.7; t = 3.6266, df = 9, p-value = 0.006). Log transformed reaction time values for the low similarity condition were correlated with identity ratings for the #1 and #2 preferred stimuli for participants (#1: t = -2.2435, df = 10, p-value = 0.049; #2: t = -2.9638, df = 10, p-value = 0.014).

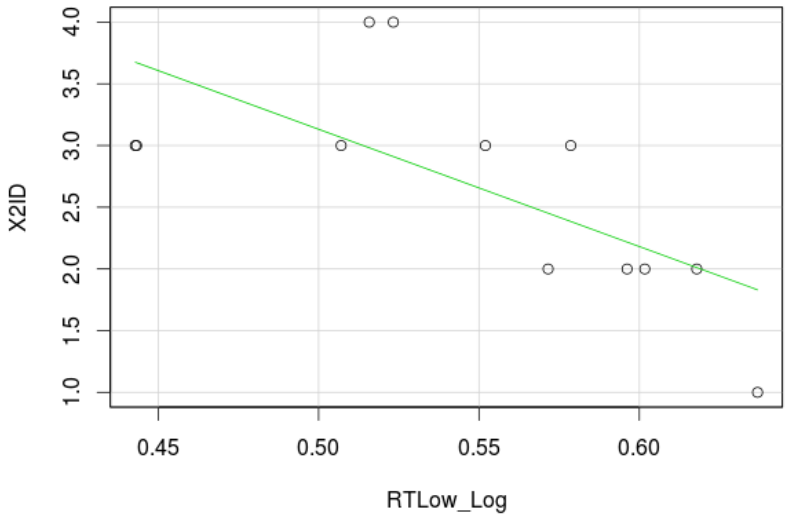


Figure 5.1. Correlation between identity ratings for the #2 preferred stimulus and log transformed reaction time during low similarity ratings.

General Linear Model

For the all similarity condition group average, activation was observed in anterior cingulate and paracingulate, bilateral putamen, accumbens, and caudate, left thalamus, bilateral insula, and left frontal pole at a cluster threshold of $z=2.33$, $p<.05$ (Figure 5.2).

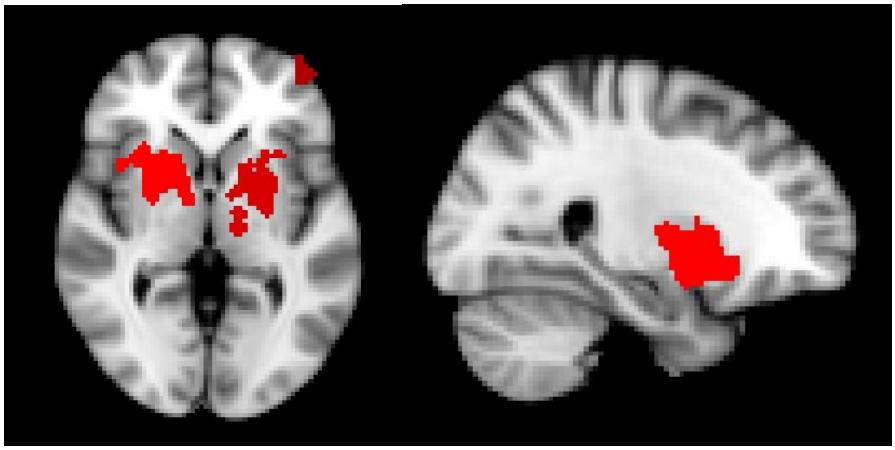


Figure 5.2. Average activation during the "all similarity" condition.

When dHPW was used as a predictor of activation during the all similarity condition, activation was observed in left amygdala, left putamen, left middle and inferior temporal gyrus, left orbitofrontal cortex, mPFC, superior frontal gyrus, anterior paracingulate and cingulate, left lateral occipital, and left temporal pole, at a cluster threshold of z=2, p<.05 (Figure 5.3).

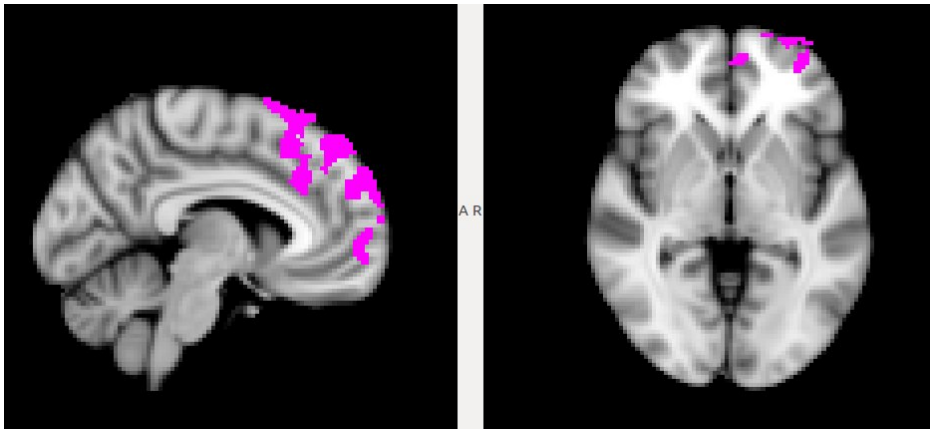


Figure 5.3. Activation during all similarity predicted by dHPW scores.

At a cluster threshold of z=1.65, p<.05, activation was predicted by aHPW in the right insula, right putamen, right planum temporale, right parietal operculum, right supramarginal gyrus, and right temporal pole (Figure 5.4).

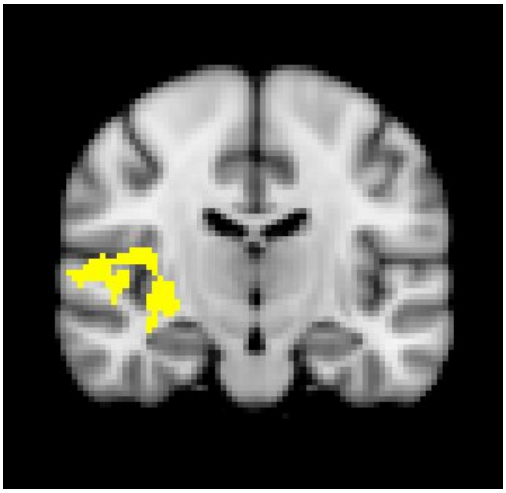


Figure 5.4. Activation in the putamen and insula predicted by aHPW ratings during the all similarity condition.

When the low similarity and high similarity conditions were compared directly, no activation was observed for the contrast of low>high above a threshold of z=1.65, p<.05. Using OLS regression, activation was observed for high>low in ventromedial and dorsomedial prefrontal regions (Figure 5.5).

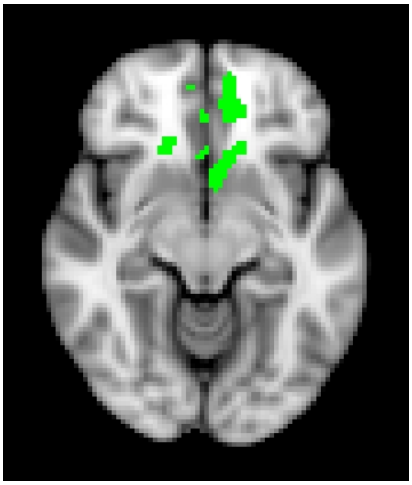


Figure 5.5. Activation greater for high similarity than low similarity ratings.

When reaction time was used as a predictor of activation during the contrast of high versus low similarity, activation was observed in bilateral OFC, VMPFC, bilateral frontal pole, right caudate, right thalamus, posterior cingulate, medial precentral gyrus, and right postcentral and supramarginal gyrus at a cluster threshold of z=1.65, p<.05 (Figure 5.6).

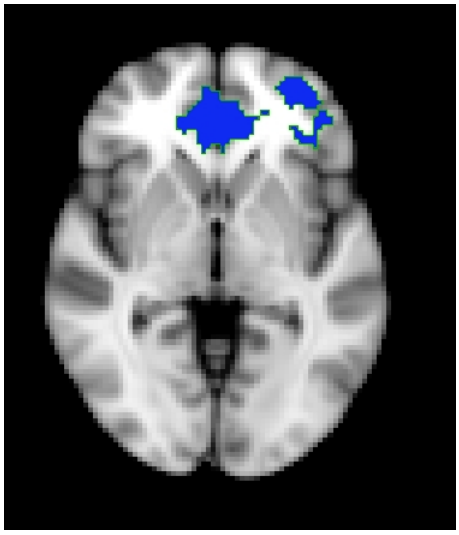


Figure 5.6. Activation predicted by RT scores that was greater during low similarity ratings compared to high.

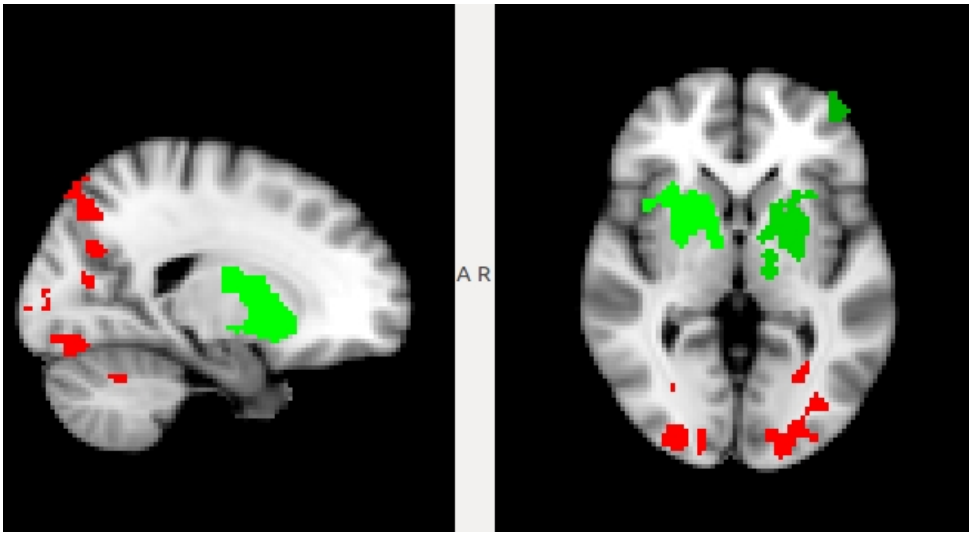


Figure 5.7. Average activation during all similarity judgements (green), and average activation during similarity judgements when subjective value and familiarity are accounted for (red).

Discussion

The goal of the present study was to characterize if and how similarity judgements for abstract reinforcers (ARs) differ from preference judgements for ARs, and if and how these similarity judgements are influenced by familiarity and subjective value. Similarity rating of ARs were related to activation in a similar cortico-striatal circuit as was involved in preference judgements for ARs, suggesting the despite very different task demands, response to these stimuli recruits similar brain networks. Our results also suggest that subjective value and familiarity with stimuli account for a significant amount of variance in brain activation during similarity judgements. These results support a bottom-up account of VBAC (value based attentional capture) during similarity ratings for abstract reinforcers.

Overall effects of similarity were observed during the task. Task activation above baseline was only observed for the contrast of high vs. low in primarily frontal regions such as

OFC and VMPFC. Reaction time was faster during high similarity judgements than low, although these differences were only statistically significant in our pilot sample (see behavioral results section). High similarity categorical judgements are generally considered more difficult than low similarity judgements (e.g. Podgorny & Garner, 1979), for example judging the difference between an eagle and a pigeon is easier than discerning the difference between a cardinal and a robin. Thus, the greater activation during high similarity judgements compared to low may reflect a difficulty effect. Greater frontal activation during low similarity>high was only observed when RT was used as a predictor of activation. These results suggest that mPFC is involved in both high and low similarity judgements, but to different degrees and in response to different factors. However, given the small sample size (n=12), more subjects may be required to clarify the relationship between frontal activation and similarity ratings. Additionally, RT scores were also correlated with self report of the relationship between most preferred stimuli and identity (Figure 5.1). Subjects who rated stimuli as less self-relevant took longer to respond, and those with the highest identity ratings responded fastest during low similarity judgements. Thus, affinity for stimuli (in the form of considering them relevant to one's self concept) may facilitate faster dissimilarity judgements. Future studies should expand the identity rating scale to better characterize individual differences. Combined with a larger sample size, this approach might be more sensitive in detecting direct relationships between identity ratings and brain activation, a relationship that was not observed in the current analysis.

Studies of similarity judgements of neutral stimuli (e.g. animals or tools) typically involve processing primarily in temporal, parietal, and occipital regions. Knowledge of such stimuli is guided by sensory and functional features learned through experience (i.e., what a hammer looks like and what it does), and this is reflected in activation in networks associated with visual processing and categorical knowledge (e.g. Kalenine et al., 2009). If participants were completing our similarity task using primarily categorical knowledge, that is, sensory features

(such as the climate in different countries, or the sensation of hiking versus biking) or functional features (reading a novel serves a different function than reading a text book), it would be expected that related regions would be preferentially recruited. Instead, the activation in the "all similarity" condition reflect a network of areas involved in decision making more generally, such as left orbitofrontal cortex, mPFC, anterior paracingulate and cingulate; visual processing, e.g. lateral occipital cortex; and also affective processing, including the left amygdala and left putamen. Areas such as anterior cingulate and caudate are also associated with the late stages of gradual object recognition leading up to an explicit decision to mark stimuli as recognized (Ploran et al., 2007). Here, these areas may process the decision related aspects of the task. Additionally, although hippocampal activation was observed at a cluster threshold of $z=1.65$, $p<.05$, this activation did not survive a cluster threshold of $z=2.33$, $p<.05$. As the hippocampus is widely implicated in memory and associative processing, this suggests that while memory may drive some of the activation during similarity ratings, the activation in striatum and frontal cortex recruit more brain resources during the task in the form of higher magnitude of activation. Finally, activation of the amygdala in particular suggest processing beyond just simple sensory or functional dimensions, as this sort of limbic circuitry is not typically active during other similarity or categorization tasks. In order to further clarify task dynamics, subject ratings of familiarity and subjective value were used as predictors of activation during the task.

Familiarity, measured as actual hours per week subjects engage with preferred stimuli, predicted activation in the insula and caudate. Subjective value, measured as the desired hours per week subjects would engage with stimuli if time or money weren't an obstacle, was related to activation in both frontal regions (mPFC, OFC, superior frontal gyrus, anterior cingulate) and limbic regions (left amygdala and putamen) during the all similarity condition. When familiarity and subjective value were used as predictors of activation during the all similarity condition, the average activation related to just similarity appears more in line with a categorical processing

network (Figure 5.7). That is, when value and familiarity are controlled for, task related activation is largely observed in occipital cortex. These results not only provide support for a bottom-up account of VBAC, they also suggest that processes related to familiarity and value more strongly drive magnitude of response during the task.

Taken together, these results suggest that similarity judgements may not be "cold" evaluations, involving brain areas primarily implicated in attentional and perceptual processing such as lateral occipital cortex, but instead recruit both frontal areas and limbic regions associated with reward value. These results also provide support for a bottom-up account of VBAC. Participants completed the task with minimal value priming - although they selected categories based on preference, they were not given any explicit instruction to consider value when making judgements. Value ratings were collected after each subject completed the task. Despite this, value ratings predict a large amount of task based activation, suggesting that it is hard for subjects to separate their preconceived value of stimuli from their similarity ratings. Any value-related behavior or brain activation is uninstructed and most likely implicit. However, it is possible some subjects decided to use explicit, value-based strategies to rate similarity (i.e. "things I like in equal amounts are similar"). Future analyses will examine evaluative dimensions used in similarity ratings to establish the extent to which value guides behavior.

These finding have important implications for the study of decision making. First, it suggests that VBAC is not an artifact of training in the lab, which may be the case when subjects are repeatedly asked for value ratings prior to completing the value-irrelevant task. Second, it suggests that for real-world relevant stimuli in particular, value to the observer plays an important part in judgements of similarity. This is both a limitation and an advantage - a limitation because it suggests these stimuli are not be processed objectively. That is, processing is not limited to object-based visual attention or categorization. Rather, value-based processing provides a means by which the interaction of valuation and other dimensions can be quantified. Recognizing that

value plays a role in similarity judgements can greatly affect how tasks engaging cognitive processes such as working memory, social interaction, or emotion regulation are studied

In sum, these results demonstrate that similarity judgements for ARs appear to invoke processes associated with reward-based decision making rather than strictly object-based categorization processes. Acknowledging the role played by VBAC during cognitive tasks allows the positing of a more comprehensive mechanism whereby features of real-world stimuli drive attention, and provides a framework for studying how value may be implicitly involved in other task scenarios.

Conclusions

The AR (abstract reinforcer) task was developed with the goal of implementing a contingency-free, multidimensional choice context in order to characterize how idiosyncratic values, attitudes, and decision making biases are represented in the brain during decision making. We manipulated a number of key factors we believe characterize the processes involved in abstract reinforcer tasks that are both shared and distinct from those involved in concrete reinforcer tasks. These factors include context (framing effects), subjective value, and valence (appetitive / aversive). In this section, results from studies 1a, 1b, 2 and 3 are discussed in terms of their key findings, with emphasis on how the results of these studies together increase understanding of the processes underlying decision making.

Framing of Appetitive vs. Aversive Reinforcers

Recent research has begun to characterize brain response to hypothetical rewards (e.g. Sharot et al., 2009; Bray et al., 2010; Miyapuram et al., 2012) and response to multimodal reward stimuli (e.g., Gross et al., 2014). The results of Studies 1-3 extend these findings by drawing on both of these approaches to investigate differences related to framing effects, task type, and valence. We show that binary choice scenarios produce very different behavior and brain activation based on the framing of choice prompts as positive or negative and the valence of stimuli being appetitive or aversive. Furthermore, we have found that framing and valence also interact to affect decisions and that these behavioral effects are mappable to brain activity.

In Study 1 we demonstrated that framing effects influence choices for appetitive abstract reinforcers. Specifically, positive frames ("which do you like more") recruit different regions of the decision making network than loss frames ("which do you like less"). Positive framing resulted in greater activation in DLPFC, an area implicated in attentional control, whereas negative framing resulted in higher magnitude of activation in VMPFC, an area implicated in coding value. Similar overall connectivity patterns were observed during both conditions, but

negative framing involved stronger connectivity amongst the anterior cingulate, insula, and striatum, suggesting that frame-based differences in response are driven by both limbic and frontal dynamics (Figure 6.1).

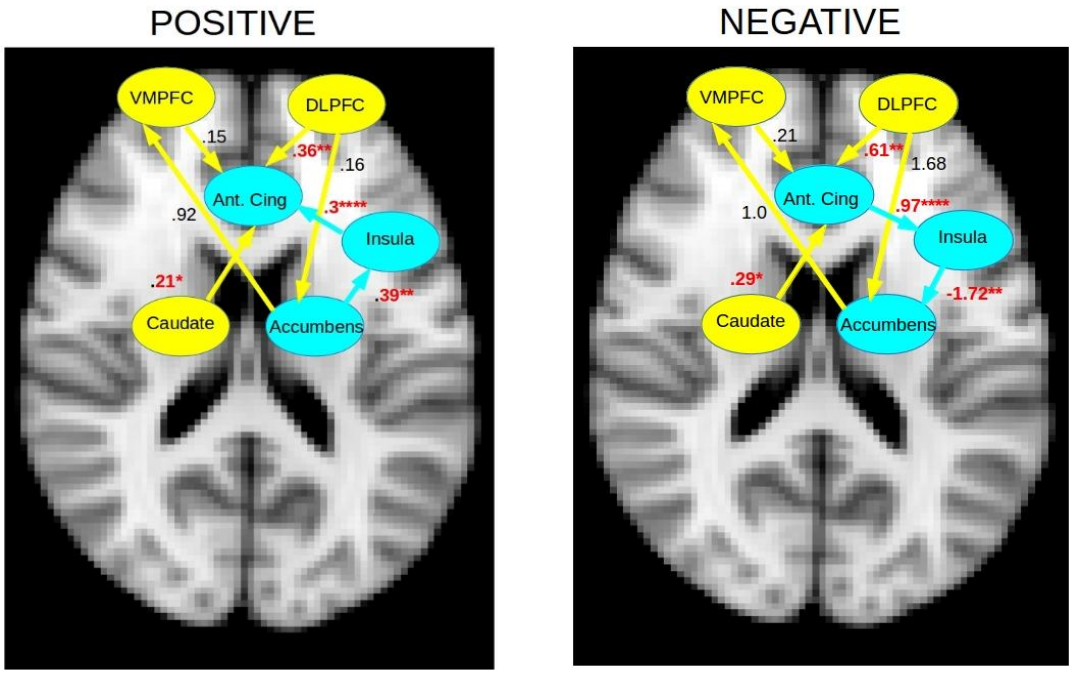


Figure 6.1.

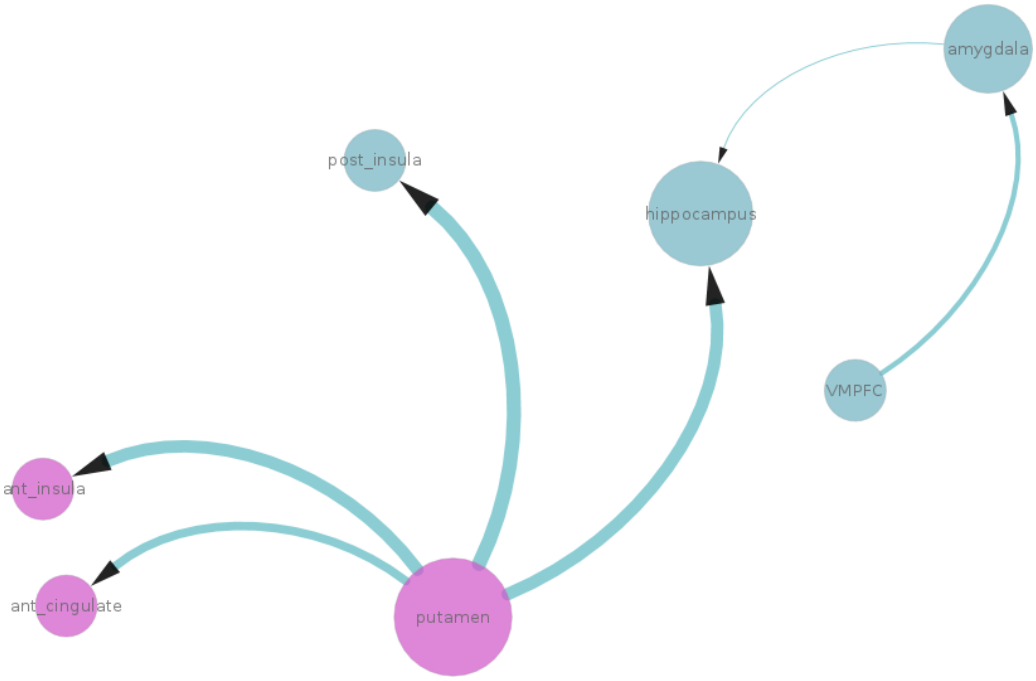
Connectivity results from Study 1a, positive ("which do you like more") versus negative ("which do you like less") framing.

Study 2 extended these findings by using aversive stimuli such as car accidents and illnesses, and framing those choices as approach or avoidance ("which would you rather have/avoid"). Smaller differences were observed in terms of changes in magnitude of activation during approach vs. avoidance compared to framing of appetitive ARs, suggesting a potential ceiling or floor effect for aversive stimuli - when all hypothetical outcomes are negative, it may matter less to the brain whether you are running away from a bear or towards a pit of snakes. Whether the effect of framing for aversive stimuli reflected a ceiling or floor effect was difficult

to determine given the limitations in current GLM analysis tools to compare patterns of deactivation in the brain.

Whereas the GLM analysis was somewhat ambiguous with regard to framing effects, the connectivity analysis clearly differentiated networks associated with approach/avoidance decisions. Specifically there was more communication between active and deactive areas as well as within deactive regions during avoidance framing compared to approach (Figure 6.2a+b). These results suggest that unlike the activation-based differences for appetitive stimuli, framing effects on aversive ARs largely occurs via differences in deactivation (see next section, "valence effects on abstract reinforcers"). This deactivation occurs in regions associated with value and affective response, although both the involved regions and the patterns of connectivity between them were different than for appetitive framing. Specifically, instead of changes in direction and strength of connections by frame as seen in the appetitive case, there was an increase in the number of connections during avoidance framing within deactive regions. Additionally, instead of anterior cingulate being central to the network, for aversive framing the putamen was central during both approach and avoidance. The putamen may code factors such as the hypothetical aversiveness of the choice options, or the intensity/perceived severity of those choices, information that may be incorporated into a value signal in value sensitive regions such as mPFC. Amygdala connectivity also changed in response to frame, with an additional connection observed from posterior insula->amygdala during avoidance framing. Given the role of the amygdala in responding to aversive stimuli (e.g. O'Doherty, 2001; Whalen et al., 2004; Orsini et al., 2015), particularly during fear learning (e.g. Nader et al., 2000; Wolff et al., 2014; Moscarello et al., 2014), these results suggest that inputs from the posterior insula may directly influence amygdala response to framing of aversive ARs, such as by relaying information about relevant sensory features.

Taken together, these results suggest differences in the mechanisms underlying framing effects on appetitive vs. aversive ARs, including activation increases for appetitive but decreases for aversive ARs; and changes in connection strength for appetitive ARs, versus number of connections for aversive ARs. These results suggest that valence of stimuli has a significant effect on brain response to framing.



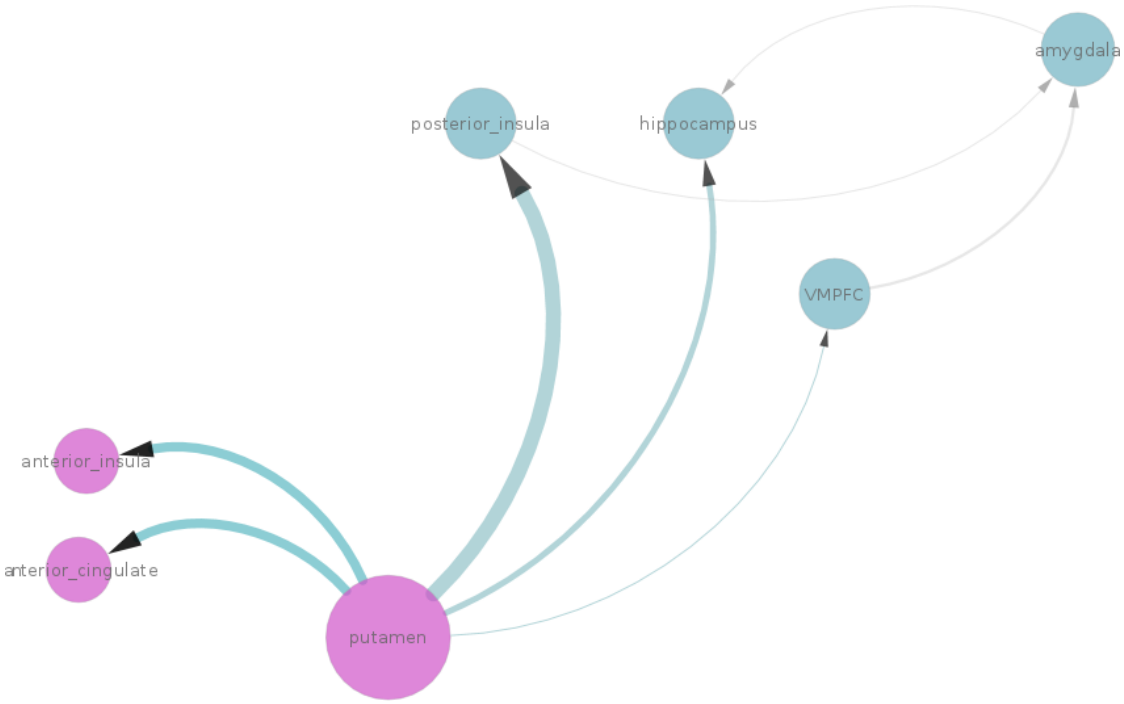


Figure 6.2. Connectivity analysis of approach framing. Node color: pink=active region, blue=deactive region. Edge thickness: thicker=higher coefficient. Arrows indicate direction of connection.

Valence effects on Abstract Reinforcers

Results from studies 1 and 2 suggest that in the appetitive domain, coding of abstract reinforcers primarily involves increases in magnitude of activation, whereas for aversive stimuli brain response primarily involves increased deactivation. This distinction is important because the salience account of decision making suggests that as stimulus intensity increases, activation magnitude should also increase. This account would predict increases in activation for both highly valuable appetitive stimuli and highly aversive stimuli. However, the valence coding account predicts activation increases for appetitive stimuli and decreases for aversive stimuli.

In order to further clarify how magnitude increases and decreases differ based on valence of stimuli, activation during the appetitive high value and aversive high severity conditions were compared directly. For all contrasts of aversive>appetitive, no activation was observed above a

cluster threshold of z=1.65, p=.05. For the contrasts of appetitive vs. aversive, activation was observed that differed by frame. Specifically, for positive appetitive framing ("which do you like more") compared to avoidance aversive framing ("which would you rather avoid"), greater activation was observed in bilateral insula, anterior and posterior cingulate, precuneus, and bilateral lateral occipital cortex at a cluster threshold of z=1.65, p=.05 (Figure 6.3). For negative appetitive framing ("which do you like less") compared to approach framing for aversive stimuli ("which would you rather have"), greater activation was observed in VMPFC at a cluster threshold of z=2.33, p=.05 (Figure 6.4).

These results provide support for valence sensitive coding during decisions about abstract reinforcers. That is, since activation for appetitive framing is greater than that for aversive framing in both framing conditions, this suggests that valence influences choices for abstract reinforcers by leading to increases when the ARs are appetitive and decreases when they are aversive. Within the concrete domain, there are results in favor of both the salience and valence account, and these differences may be due to protocol differences, such as contextual changes (gambling vs. certain choices, learning vs. passive tasks, etc.) that may drive responding to be more activation or deactivation based. Here, the protocols were nearly identical and subjects were scanned using the same scanner, so cross-protocol differences are as minimal as possible. However, future analyses could use measures such as percent signal change to characterize the activation increases and decreases in each condition, in particular to determine if areas such as mPFC increase or decrease activation in a manner that is parametrically related to increase and decreases in stimulus value.

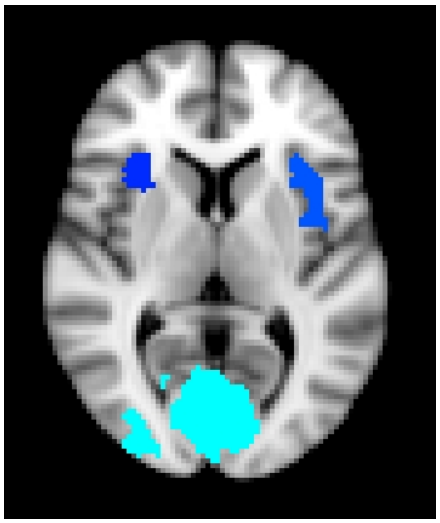


Figure 6.3. Activation that was greater for appetitive positive framing than aversive avoidance framing.

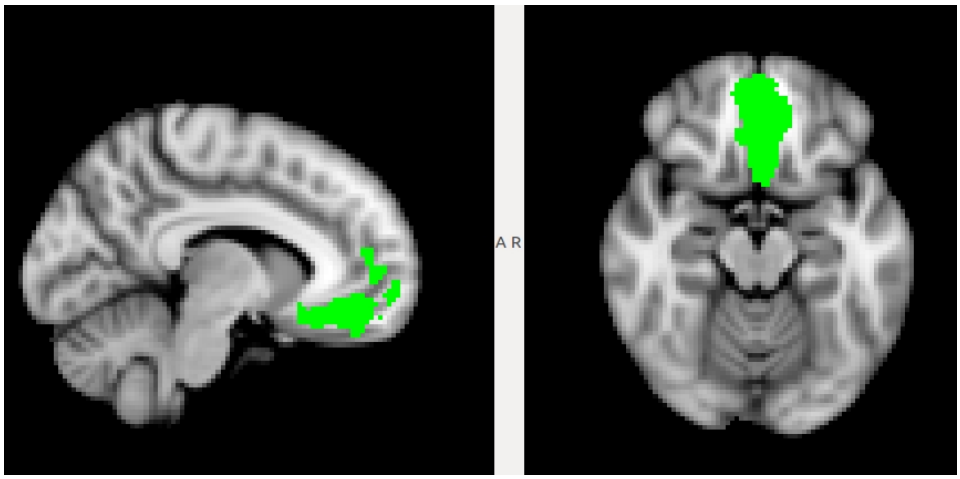


Figure 6.4. Activation that was greater for appetitive negative framing than aversive approach framing.

Familiarity effects on choices for abstract reinforcers

Familiarity with stimuli is a significant factor that can influences choice. Simply being passively exposed to a stimulus artificially increases liking of that stimulus (Zajonc, 1968). In the aversive domain, choosing one hypothetical illness over another decreases the perceived aversiveness of the chosen illness, suggesting that mere exposure can lead to decreases in dislike as well (Sharot et al, 2010). Thus, in Studies 1-3 familiarity with stimuli was measured and used

as a predictor of brain activation in order to determine the extent of influence that familiarity has on abstract reinforcer decisions.

Familiarity effects differed significantly based on task type and stimulus valence. Familiarity was measured in Studies 1a+b and 3 using actual hours per week (aHPW), the average hours subjects engaged with their top three preferred stimuli (see 1b Methods section for rationale). During preference judgements for appetitive ARs in Study 1a+b, familiarity ratings did not predict any brain activation above threshold, or correlate with any behavioral variables. This is important in showing that the observed brain activation and behavior were due to more than mere exposure effects inasmuch as familiarity would be expected to predict activation. However, the lack of influence related to familiarity must be interpreted with caution due to the risk of Type II error.

Familiarity scores predicted activation during similarity ratings of ARs in Study 3, and during choices for aversive stimuli in Study 2. In Study 2, familiarity was rated on a scale from 1 to 7 (1=totally unfamiliar, 7=highly familiar) for each stimulus, and each subject's average rating was used as a predictor of activation during the task. During the similarity task, aHPW scores were related to activation in the insula and caudate. These regions are implicated in coding stimulus intensity and reward value. These results suggest that the effect of familiarity, as well as that of value based attentional capture (VBAC), influences brain response during similarity ratings. During choices for aversive ARs, aHPW predicted activation in orbitofrontal cortex and parahippocampus, suggesting their involvement in representing familiar stimuli. However, aHPW did not predict any activation during appetitive choices, suggesting that familiarity plays a different role during choices for aversive stimuli and similarity ratings than during preference judgements. Future studies could have participants provide ratings of the familiarity, aHPW and dHPW for every stimulus in order to determine how familiarity interacts valence and choice context.

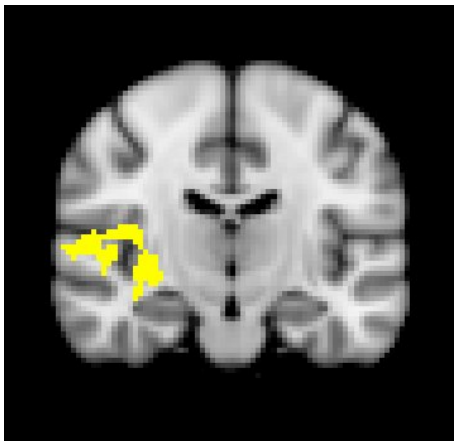


Figure 6.5. Activation predicted by familiarity score during similarity ratings

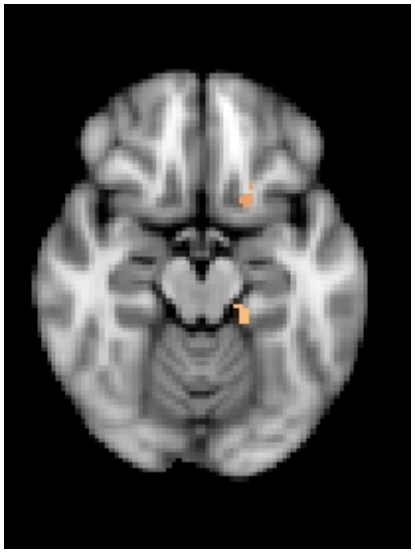


Figure 6.6. Right, activation predicted by familiarity during aversive choices (orange).

Valuation of abstract reinforcers

The goal of Study 1b was to determine if the subjective value of appetitive ARs is coded in the brain in a way comparable to that used in the value coding of concrete reinforcers. A

similar methodology was used in Study 3 to determine how subjective value influences choice behavior implicitly via value based attentional capture (VBAC).

Study 1b tested the "common neural currency" hypothesis of value coding using ARs. According to this hypothesis, even when choosing amongst stimuli that cannot be evaluated on an objective common scale, the brain is able to code stimuli based a common neural scaling in order to facilitate choice. For example, we can trade time for money even though time exists on a different scale than currency. Common coding of lower dimensionality stimuli such as money, food, and time has been well established using concrete reinforcers. Results from Study 1b demonstrate that even when participants choose from entirely different, highly multidimensional stimulus categories, greater activation for subjective high value vs. low value choices is observed in mPFC, thought to be the brain area that implements the common neural currency code (e.g. Chib et al., 2009). Study 1b also tested a novel quantification of subjective value for ARs - the desired hours per week (dHPW) participants would spend engaging with preferred stimuli if time or money were no obstacle. Desired, but not actual HPW (a measure of familiarity) predicted activation in mPFC during high value choices, in voxels that overlapped with those more active during the contrast of high value vs. low value.

Study 3 leveraged the AR approach developed in Study 2 to clarify how value processes interact with other choice constructs, specifically similarity ratings. Theories of attentional capture suggest that task-irrelevant factors often influence choices, but the mechanisms by which this can happen are heavily debated. Some studies suggest that attentional capture is subject to top-down control, via attentional and working memory mechanisms (Kiss et al., 2012; Eimer & Kiss, 2008; Leber & Egeth, 2006; Lamy et al., 2014), whereas others suggest it is a bottom-up, implicit process (Anderson et al., 2011a,b; Greuschow et al., 2015). Many studies test this effect in the perceptual domain (i.e. using pop-out effects), but fewer have studied how cognitive constructs such as value capture attention during decision making, a phenomenon called value

based attentional capture or VBAC (e.g. Greuschow et al., 2015). In Study 3, we find that dHPW ratings of subjective value predict activation during similarity ratings, supporting a bottom-up account of VBAC during abstract reinforcer judgements. The area of mPFC active for this analysis partially overlapped with the voxels that were related to dHPW during preference judgements, suggesting that dHPW is a reliable predictor of mPFC activity across protocols and groups (Figure 6.7).

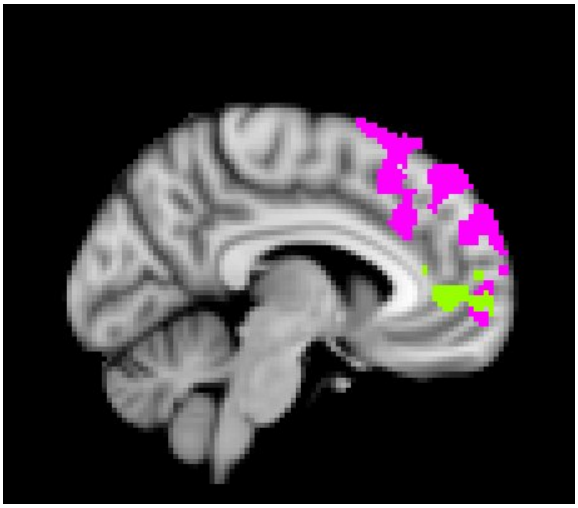


Figure 6.7. Activation predicted by dHPW during similarity ratings (pink) and preference judgements (green).

These results from Study 1a and Study 3 suggest that subjective value of ARs can be measured using dHPW, a novel approach analogous to other established subjective value measurements like willingness to pay. This is important because previous studies often quantify subjective value for real-world relevant stimuli using time or money as proxies. However, these proxy variables themselves have their own subjective value, in that money or time may have different value for different individuals. Obtaining subjective value for ARs directly from subjects removes the issue of assumed value. Furthermore, the results from Study 3 suggest that dHPW can be used to study implicit value processes that occur during other cognitive tasks,

providing a framework for studying how value may be implicitly involved in other task contexts such as working memory or attention.

An important limitation of comparisons across studies 1 and 2 is that in Study 1, participants were able to provide subjective value ratings for stimuli because the categories drawn from hobbies, activities, or interests that have value to them in real life. In Study 2, however, participants were unlikely to have equally well defined value structures, such as the top ten illnesses they would prefer to have (or avoid). In fact, this lack of a preexisting value structure for aversive stimuli may underlie what appears to be a learning effect observed for a series of choices between aversive (but not appetitive) stimuli. Therefore, although results from Study 1b indicate that activation in mPFC in particular increases in direct proportion to the value of appetitive ARs, a lack of an equivalent subjective value structure for aversive stimuli prevented an analysis of the effect of aversive ARs on areas such as mPFC and striatum. Future studies could remedy this by having participants rate the perceived aversiveness of each stimulus on a scale ranging from mildly aversive to severely aversive, to extract a ranking or scoring of stimuli which could be used to predict brain activation and thereby clarify which areas specifically code value of negatively valenced stimuli.

Future Directions

Studies 1-3 demonstrate that abstract reinforcers can be used to uniquely characterize factors that influence choices for real world relevant stimuli. This approach has broad potential applications. For example, another concrete choice context that has been well characterized is delay discounting, which quantifies individual differences in trade offs between money and time (i.e. “$1 today versus $2 tomorrow”). ARs could be used to determine if the preference functions that apply to delayed choices for money or goods also apply for stimuli like preferred hobbies. This could be done by presenting subjects with trade offs between time and ARs, like choosing between “your #3 most preferred stimulus now, or your #1 most preferred stimulus in 1 week.”

This topic is relevant in particular for understanding how to optimize trade offs like work-life balance.

Preferences for ARs are also relevant in social situations. People who share a love of running or museums often form relationships based on these shared interests. ARs may serve as a form of "social currency" that increases desirability of others as potential friends or romantic partners. This could in turn lead to biases where others who share interests are viewed more favorably than those who don't, contributing to in-group and out-group biases. These topics could be explored through a task where participants read biographies of others who either share or don't share their interests. Then they could engage in social tasks such as cyberball or trust games, and the effect of shared preferences could be quantified to determine how perception of shared interests influences behavior and brain activation.

Conclusions

The studies presented herein provide new information about brain dynamics underlying decision making in several ways. First, we considered the possibility that choices for abstract reinforcers might recruit brain regions primarily involved in future planning and mental simulation (such as DLPFC), rather than areas typically associated with decisions for concrete reinforcers (such as the insula and striatum). Furthermore, since all subjects were choosing amongst different stimuli sets, it was possible that this heterogeneity would lead to uninterpretable results at the group level (i.e. each subject having a different pattern of activation). We hypothesized instead that brain areas involved in decision making are flexible enough in their functionality that they can accommodate a wide range of choices, not just for tangible goods like money or food hypothetical abstract reinforcer choices. Due to this flexibility, we expected that group-level activation would involve overlapping patterns in brain regions associated with decision making across subjects. We find support for the functional flexibility account for choices in both the appetitive and the aversive domain, specifically recruitment of

brain areas widely associated with decision making including striatum, insula, anterior cingulate, and dorsolateral and medial prefrontal cortex. Next, we tested whether brain response to abstract reinforcers is sensitive to valence or salience, accounts which make different predictions about the magnitude-based mechanisms underlying stimulus processing. In support of a valence sensitivity account, activation during choices for aversive ARs involved both increases and decreases in activation. Finally, we tested whether subjective value of ARs implicitly influences stimulus judgements that are value-irrelevant (similarity ratings), and found that they do. This is important because it suggests that subjective preferences influence brain activation even when preference is not relevant, a bias that may extend to other cognitive functions such as memory encoding or attention.

The broadest implications of this work are related to how value and reward are defined in the field of decision making. Defining value as monetary worth ignores aspects of subjective value that are not easily quantified using money, such as the value of a social relationship. Defining a reward or reinforcer as a stimulus that reinforces a behavior similarly ignores rewards that do not linearly relate to behavioral change, such as the intrinsic reward (or cost)of daydreaming.

While these definitions have been gradually expanding, the studies presented here suggest that reward and value processes occur for a much wider range of stimuli than previously measured using neuroimaging. The human brain seems uniquely able to process nearly any stimulus subjectively valued as positive as a reward, including intrinsic rewards like choosing things one likes merely for the sake of choosing, rather than in pursuit of a goal or outcome. It appears that when primary needs like food, money and shelter are met and we pursue other needs such as leisure or play, the same circuitry that orients us to the tastiest food or the best gamble also supports the pursuit of these higher order needs. This opens up a wide range of questions for

study, such as how preferences for abstract reinforcers are formed, how these preferences evolve over time, and how the brain supports such processes.

References

Alós-Ferrer, C., Granić, Đ.-G., Shi, F., & Wagner, A. K. (2012). Choices and preferences: Evidence from implicit choices and response times. *Journal of Experimental Social Psychology*, *48*(6), 1336–1342. doi:10.1016/j.jesp.2012.07.004

Anderson, B. a., Laurent, P. a., & Yantis, S. (2011). Learned value magnifies salience-based attentional capture. *PLoS ONE*, *6*(11). doi:10.1371/journal.pone.0027926

Anderson, B. a, Laurent, P. a, & Yantis, S. (2011). Value-driven attentional capture. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(25), 10367–10371. doi:10.1073/pnas.1104047108

Ballard, K., & Knutson, B. (2009). Dissociable neural representations of future reward magnitude and delay during temporal discounting. *NeuroImage*, *45*(1), 143–50. doi:10.1016/j.neuroimage.2008.11.004

Barron, H. C., Dolan, R. J., & Behrens, T. E. J. (2013). Online evaluation of novel choices by simultaneous representation of multiple memories. *Nature neuroscience*, *16*(10), 1492–8. doi:10.1038/nn.3515

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*, 412–27. doi:10.1016/j.neuroimage.2013.02.063

Bechara, a, Damasio, H., & Damasio, a R. (2000). Emotion, decision making and the orbitofrontal cortex. *Cerebral cortex (New York, N.Y. : 1991)*, *10*(3), 295–307. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10731224

Benoit, R. G., Szpunar, K. K., & Schacter, D. L. (2014). Ventromedial prefrontal cortex supports affective future simulation by integrating distributed knowledge. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(46), 16550–5. doi:10.1073/pnas.1419274111

Bhanji, J. P., & Delgado, M. R. (2014). Perceived control influences neural responses to setbacks and promotes persistence. *Neuron*, *83*(6), 1369–75. doi:10.1016/j.neuron.2014.08.012

Bickel, W. K., Pitcock, J. a, Yi, R., & Angtuaco, E. J. C. (2009). Congruence of BOLD response across intertemporal choice conditions: fictive and real money gains and losses. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *29*(27), 8839–46. doi:10.1523/JNEUROSCI.5319-08.2009

Bickel, W. K., Yi, R., Landes, R. D., Hill, P. F., & Baxter, C. (2011). Remember the future: working memory training decreases delay discounting among stimulant addicts. *Biological psychiatry*, *69*(3), 260–5. doi:10.1016/j.biopsych.2010.08.017

Bissonette, G. B., Gentry, R. N., Padmala, S., Pessoa, L., & Roesch, M. R. (2014). Impact of appetitive and aversive outcomes on brain responses: linking the animal and human literatures. *Frontiers in systems neuroscience*, *8*(March), 24. doi:10.3389/fnsys.2014.00024

Biswas, D., & Grau, S. L. (n.d.). Product Option Framing : Loss Aversion Principles or Sensitivity to Price Differentials ?, *25*(May 2008), 399–415. doi:10.1002/mar

Blood, A. J., Zatorre, R. J., Bermudez, P., & Evans, A. C. (1999). Emotional responses to pleasant and unpleasant music correlate with activity in paralimbic brain regions, *2*(4).

Blood, A., & Zatorre, R. (2001). Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. *... of the National Academy of Sciences*, *98*(20), 11818–11823. doi:10.1073/pnas.

Boorman, E. D., Behrens, T. E. J., Woolrich, M. W., & Rushworth, M. F. S. (2009). How green is the grass on the other side? Frontopolar cortex and the evidence in favor of alternative courses of action. *Neuron*, *62*(5), 733–43. doi:10.1016/j.neuron.2009.05.014

Bouret, S., & Richmond, B. J. (2010). Ventromedial and orbital prefrontal neurons differentially encode internally and externally driven motivational values in monkeys. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *30*(25), 8591–601. doi:10.1523/JNEUROSCI.0049-10.2010

Bray, S., Shimojo, S., & O'Doherty, J. P. (2010). Human medial orbitofrontal cortex is recruited during experience of imagined and real rewards. *Journal of neurophysiology*, *103*(5), 2506–12. doi:10.1152/jn.01030.2009

Breiter, H. C., Aharon, I., Kahneman, D., Dale, a, & Shizgal, P. (2001). Functional imaging of neural responses to expectancy and experience of monetary gains and losses. *Neuron*, *30*(2), 619–39. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11395019

Bromberg-Martin, E. S., Matsumoto, M., & Hikosaka, O. (2010). Dopamine in motivational control: rewarding, aversive, and alerting. *Neuron*, *68*(5), 815–34. doi:10.1016/j.neuron.2010.11.022

Brooks, A. M., & Berns, G. S. (2013). Aversive stimuli and loss in the mesocorticolimbic dopamine system. *Trends in Cognitive Sciences*, *17*(6), 281–286. doi:10.1016/j.tics.2013.04.001

Brosch, T., Coppin, G., Schwartz, S., & Sander, D. (2012). The importance of actions and the worth of an object: dissociable neural systems representing core value and economic value. *Social cognitive and affective neuroscience*, *7*(5), 497–505. doi:10.1093/scan/nsr036

Caria, A., Sitaram, R., Veit, R., Begliomini, C., & Birbaumer, N. (2010). Volitional control of anterior insula activity modulates the response to aversive stimuli. A real-time functional magnetic resonance imaging study. *Biological psychiatry*, *68*(5), 425–32. doi:10.1016/j.biopsych.2010.04.020

Cavallo, A., Heyes, C., Becchio, C., Bird, G., & Catmur, C. (2014). Timecourse of mirror and counter-mirror effects measured with transcranial magnetic stimulation. *Social Cognitive and Affective Neuroscience*, *9*, 1082–1088. doi:10.1093/scan/nsu085

Charest, I., Kievit, R. a., Schmitz, T. W., Deca, D., & Kriegeskorte, N. (2014). Unique semantic space in the brain of each beholder predicts perceived similarity. *Proceedings of the National Academy of Sciences*, *111*(40), 14565–14570. doi:10.1073/pnas.1402594111

Chaudhry, A. M., Parkinson, J. a, Hinton, E. C., Owen, A. M., & Roberts, A. C. (2009). Preference judgements involve a network of structures within frontal, cingulate and insula cortices. *The European journal of neuroscience*, *29*(5), 1047–55. doi:10.1111/j.1460-9568.2009.06646.x

Chib, V. S., Rangel, A., Shimojo, S., & O'Doherty, J. P. (2009). Evidence for a common representation of decision values for dissimilar goods in human ventromedial prefrontal cortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *29*(39), 12315–20. doi:10.1523/JNEUROSCI.2575-09.2009

Chikazoe, J., Lee, D. H., Kriegeskorte, N., & Anderson, A. K. (2014). Population coding of affect across stimuli, modalities and individuals. *Nature neuroscience*, *17*(8), 1114–1122. doi:10.1038/nn.3749

Chiu, Y., Cools, R., & Aron, A. (2014). Opposing Effects of Appetitive and Aversive Cues on Go/No-go Behavior and Motor Excitability, 1–10. doi:10.1162/jocn

Clithero, J. a, & Rangel, A. (2014). Informatic parcellation of the network involved in the computation of subjective value. *Social cognitive and affective neuroscience*, *9*(9), 1289–302. doi:10.1093/scan/nst106

Cohen, M. X., Heller, a S., & Ranganath, C. (2005). Functional connectivity with anterior cingulate and orbitofrontal cortices during decision-making. *Brain research. Cognitive brain research*, *23*(1), 61–70. doi:10.1016/j.cogbrainres.2005.01.010

Collins, K. a., Mendelsohn, A., Cain, C. K., & Schiller, D. (2014). Taking Action in the Face of Threat: Neural Synchronization Predicts Adaptive Coping. *Journal of Neuroscience*, *34*(44), 14733–14738. doi:10.1523/JNEUROSCI.2152-14.2014

Cunningham, W. a., & Brosch, T. (2012). Motivational Salience: Amygdala Tuning From Traits, Needs, Values, and Goals. *Current Directions in Psychological Science*, *21*(1), 54–59. doi:10.1177/0963721411430832

De Martino, B., Kumaran, D., Seymour, B., & Dolan, R. J. (2006). Frames, biases, and rational decision-making in the human brain. *Science (New York, N.Y.)*, *313*(5787), 684–7. doi:10.1126/science.1128356

Delgado, M R, Locke, H. M., Stenger, V. a, & Fiez, J. a. (2003). Dorsal striatum responses to reward and punishment: effects of valence and magnitude manipulations. *Cognitive, affective & behavioral neuroscience*, *3*(1), 27–38. doi:10.3758/CABN.3.1.27

Delgado, Mauricio R, Li, J., Schiller, D., & Phelps, E. a. (2008). The role of the striatum in aversive learning and aversive prediction errors. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *363*(1511), 3787–800. doi:10.1098/rstb.2008.0161

Dellacherie, D., Bigand, E., Molin, P., Baulac, M., & Samson, S. (2011). Multidimensional scaling of emotional responses to music in patients with temporal lobe resection. *Cortex; a journal devoted to the study of the nervous system and behavior*, *47*(9), 1107–15. doi:10.1016/j.cortex.2011.05.007

Díaz-Santos, M., Cao, B., Yazdanbakhsh, A., Norton, D. J., Neargarder, S., & Cronin-Golomb, A. (2015). Perceptual, cognitive, and personality rigidity in Parkinson's disease. *Neuropsychologia*, *69*, 183–193. doi:10.1016/j.neuropsychologia.2015.01.044

D'Argembeau, A., Xue, G., Lu, Z.-L., Van der Linden, M., & Bechara, A. (2008). Neural correlates of envisioning emotional events in the near and far future. *NeuroImage*, *40*(1), 398–407. doi:10.1016/j.neuroimage.2007.11.025

Edelman, S., Grill-Spector, K., Kushnir, T., & Malach, R. (1998). Toward direct visualization of the internal shape represetation space by fMRI. *Psychobiology*, *26*(4), 309–321.

Eimer, M., & Kiss, M. (2008). Involuntary attentional capture is determined by task set: evidence from event-related brain potentials. *Journal of cognitive neuroscience*, *20*, 1423–1433. doi:10.1162/jocn.2008.20099

Etkin, A., Egner, T., & Kalisch, R. (2011). Emotional processing in anterior cingulate and medial prefrontal cortex. *Trends in cognitive sciences*, *15*(2), 85–93. doi:10.1016/j.tics.2010.11.004.Emotional

Feldman Hall, O., Dalgleish, T., Thompson, R., Evans, D., Schweizer, S., & Mobbs, D. (2012). Differential neural circuitry and self-interest in real vs hypothetical moral decisions. *Social Cognitive and Affective Neuroscience*, *7*, 743–751. doi:10.1093/scan/nss069

FeldmanHall, O., Dalgleish, T., Thompson, R., Evans, D., Schweizer, S., & Mobbs, D. (2012). Differential neural circuitry and self-interest in real vs hypothetical moral decisions. *Social cognitive and affective neuroscience*, *7*(7), 743–51. doi:10.1093/scan/nss069

Ferguson, S. G., & Shiffman, S. (2009). The relevance and treatment of cue-induced cravings in tobacco dependence. *Journal of substance abuse treatment*, *36*(3), 235–43. doi:10.1016/j.jsat.2008.06.005

Flexas, A., Rosselló, J., de Miguel, P., Nadal, M., & Munar, E. (2014). Cognitive control and unusual decisions about beauty: an fMRI study. *Frontiers in human neuroscience*, *8*(July), 520. doi:10.3389/fnhum.2014.00520

Foo, J. C., Haji, T., & Sakai, K. (2014). Prefrontal Mechanisms in Preference and Non-Preference-based Judgments. *NeuroImage*. doi:10.1016/j.neuroimage.2014.03.046

Furnham, A., & Boo, H. C. (2011). A literature review of the anchoring effect. *Journal of Socio-Economics*, *40*(1), 35–42. doi:10.1016/j.socec.2010.10.008

Garvert, M. M., Moutoussis, M., Behrens, T. E. J., & Dolan, R. J. (2015). Learning-Induced Plasticity in Medial Prefrontal Article Learning-Induced Plasticity in Medial Prefrontal Cortex Predicts Preference Malleability. *Neuron*, *85*(2), 418–428. doi:10.1016/j.neuron.2014.12.033

Gasquoine, P. G. (2013). Localization of function in anterior cingulate cortex: from psychosurgery to functional neuroimaging. *Neuroscience and biobehavioral reviews*, *37*(3), 340–8. doi:10.1016/j.neubiorev.2013.01.002

Gilaie-Dotan, S., Tymula, a., Cooper, N., Kable, J. W., Glimcher, P. W., & Levy, I. (2014). Neuroanatomy Predicts Individual Risk Attitudes. *Journal of Neuroscience*, *34*(37), 12394–12401. doi:10.1523/JNEUROSCI.1600-14.2014

Gläscher, J., Hampton, A. N., & O'Doherty, J. P. (2012). Determining a role for ventromedial prefrontal cortex in encoding action-based value signals during reward-related decision making. *Cerebral cortex (New York, N.Y. : 1991)*, *19*(2), 483–95. doi:10.1093/cercor/bhn098

Gold, B. P., Frank, M. J., Bogert, B., & Brattico, E. (2013). Pleasurable music affects reinforcement learning according to the listener. *Frontiers in psychology*, *4*(August), 541. doi:10.3389/fpsyg.2013.00541

Gonzalez, C., Dana, J., Koshino, H., & Just, M. (2005). The framing effect and risky decisions: Examining cognitive functions with fMRI. *Journal of Economic Psychology*, *26*(1), 1–20. doi:10.1016/j.joep.2004.08.004

Gottfried, J. a, O'Doherty, J., & Dolan, R. J. (2002). Appetitive and aversive olfactory learning in humans studied using event-related functional magnetic resonance imaging. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *22*(24), 10829–37. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12486176

Grabenhorst, F., & Rolls, E. T. (2011). Value, pleasure and choice in the ventral prefrontal cortex. *Trends in cognitive sciences*, *15*(2), 56–67. doi:10.1016/j.tics.2010.12.004

Gross, J., Woelbert, E., Zimmermann, J., Okamoto-Barth, S., Riedl, A., & Goebel, R. (2014). Value signals in the prefrontal cortex predict individual preferences across reward categories. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *34*(22), 7580–6. doi:10.1523/JNEUROSCI.5082-13.2014

Grueschow, M., Polania, R., Hare, T. A., & Ruff, C. C. (2015). Automatic versus Choice-Dependent Value Representations in the Human Brain. *Neuron*, *85*(4), 874–885. doi:10.1016/j.neuron.2014.12.054

Gu, X., Liu, X., Van Dam, N. T., Hof, P. R., & Fan, J. (2013). Cognition-emotion integration in the anterior insular cortex. *Cerebral cortex (New York, N.Y. : 1991)*, *23*(1), 20–7. doi:10.1093/cercor/bhr367

Guitart-Masip, M., Bunzeck, N., Stephan, K. E., Dolan, R. J., & Düzel, E. (2010). Contextual novelty changes reward representations in the striatum. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *30*(5), 1721–6. doi:10.1523/JNEUROSCI.5331-09.2010

Gusnard, D. a, Akbudak, E., Shulman, G. L., & Raichle, M. E. (2001). Medial prefrontal cortex and self-referential mental activity: relation to a default mode of brain function. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(15), 4259–4264. doi:10.1073/pnas.071043098

Hanson, C., Hanson, S. J., Ramsey, J., & Glymour, C. (2013). Atypical effective connectivity of social brain networks in individuals with autism. *Brain connectivity*, *3*(6), 578–89. doi:10.1089/brain.2013.0161

Hare, T. a, Camerer, C. F., & Rangel, A. (2009). Self-control in decision-making involves modulation of the vmPFC valuation system. *Science (New York, N.Y.)*, *324*(5927), 646–8. doi:10.1126/science.1168450

Hariri, A. R., Brown, S. M., Williamson, D. E., Flory, J. D., de Wit, H., & Manuck, S. B. (2006). Preference for immediate over delayed rewards is associated with magnitude of ventral striatal activity. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *26*(51), 13213–7. doi:10.1523/JNEUROSCI.3446-06.2006

Harris, A., Adolphs, R., Camerer, C., & Rangel, A. (2011). Dynamic construction of stimulus values in the ventromedial prefrontal cortex. *PloS one*, *6*(6), e21074. doi:10.1371/journal.pone.0021074

Hayes, D. J., Duncan, N. W., Xu, J., & Northoff, G. (2014). A comparison of neural responses to appetitive and aversive stimuli in humans and other mammals. *Neuroscience and biobehavioral reviews*, *45*, 350–68. doi:10.1016/j.neubiorev.2014.06.018

Hayes, D. J., & Northoff, G. (2011). Identifying a network of brain regions involved in aversion-related processing: a cross-species translational investigation. *Frontiers in integrative neuroscience*, *5*(October), 49. doi:10.3389/fnint.2011.00049

Hayes, D. J., & Northoff, G. (2012). Common brain activations for painful and non-painful aversive stimuli. *BMC neuroscience*, *13*, 60. doi:10.1186/1471-2202-13-60

Iyengar, S. S., & Lepper, M. R. (2000). When choice is demotivating: can one desire too much of a good thing? *Journal of personality and social psychology*, *79*(6), 995–1006. doi:10.1037/0022-3514.79.6.995

J. Silvers, T. Wager, J Weber, K. O. (2014). The neural basis of uninstructed negative emotion regulation. *Social cognitive and affective neuroscience*.

Jackson;, D., & Messick, S. (1963). Individual differences in social facilitation. *British Journal of Social Clinical Psychology*, *2*, 1–10. Retrieved from http://mss3.libraries.rutgers.edu/dlr/showfed.php?pid=rutgers-lib:26428

Jarcho, J. M. J., Berkman, E. T., & Lieberman, M. D. (2010). The neural basis of rationalization: cognitive dissonance reduction during decision-making. *Social cognitive and affective neuroscience*, *6*(4), 460–7. doi:10.1093/scan/nsq054

Jaworska, N., & Chupetlovska-Anastasova, A. (2009). A review of multidimensional Scaling (MDS) and its utility in various psychological domains. *Tutorials in Quantitative Methods for Psychology*, *5*(1), 1–10. doi:10.1177/014662168300700404

Johnson, S. C., Schmitz, T. W., Kawahara-Baccus, T. N., Rowley, H. a, Alexander, A. L., Lee, J., & Davidson, R. J. (2005). The cerebral response during subjective choice with and without self-reference. *Journal of cognitive neuroscience*, *17*(12), 1897–906. doi:10.1162/089892905775008607

Joshua, M., Adler, A., Mitelman, R., Vaadia, E., & Bergman, H. (2008). Midbrain dopaminergic neurons and striatal cholinergic interneurons encode the difference between reward and aversive events at different epochs of probabilistic classical conditioning trials. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *28*(45), 11673–84. doi:10.1523/JNEUROSCI.3839-08.2008

Kahnt, T., Park, S. Q., Haynes, J.-D., & Tobler, P. N. (2014). Disentangling neural representations of value and salience in the human brain. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(13), 5000–5. doi:10.1073/pnas.1320189111

Kahnt, T., & Tobler, P. N. (2013). Salience signals in the right temporoparietal junction facilitate value-based decisions. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *33*(3), 863–9. doi:10.1523/JNEUROSCI.3531-12.2013

Kang, M. J., & Camerer, C. F. (2013). fMRI evidence of a hot-cold empathy gap in hypothetical and real aversive choices. *Frontiers in neuroscience*, *7*(June), 104. doi:10.3389/fnins.2013.00104

Kim, H., Adolphs, R., O'Doherty, J. P., & Shimojo, S. (2007). Temporal isolation of neural processes underlying face preference decisions. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(46), 18253–8. doi:10.1073/pnas.0703101104

Kim, H., Shimojo, S., & O'Doherty, J. P. (2006). Is avoiding an aversive outcome rewarding? Neural substrates of avoidance learning in the human brain. *PLoS biology*, *4*(8), e233. doi:10.1371/journal.pbio.0040233

Kim, K., & Johnson, M. K. (2014). Activity in ventromedial prefrontal cortex during self-related processing: Positive subjective value or personal significance? *Social cognitive and affective neuroscience*. doi:10.1093/scan/nsu078

Kiss, M., Grubert, A., Petersen, A., & Eimer, M. (2012). Attentional Capture by Salient Distractors during Visual Search Is Determined by Temporal Task Demands. *Journal of Cognitive Neuroscience*, *24*, 749–759. doi:10.1162/jocn_a_00127

Knutson, B., Fong, G. W., Bennett, S. M., Adams, C. M., & Hommer, D. (2003). A region of mesial prefrontal cortex tracks monetarily rewarding outcomes: characterization with rapid event-related fMRI. *NeuroImage*, *18*(2), 263–272. doi:10.1016/S1053-8119(02)00057-5

Knutson, B., Wimmer, G. E., Rick, S., Hollon, N. G., Prelec, D., & Loewenstein, G. (2008). Neural antecedents of the endowment effect. *Neuron*, *58*(5), 814–22. doi:10.1016/j.neuron.2008.05.018

Koelsch, S., Fritz, T., V Cramon, D. Y., Müller, K., & Friederici, A. D. (2006). Investigating emotion with music: an fMRI study. *Human brain mapping*, *27*(3), 239–50. doi:10.1002/hbm.20180

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. *Frontiers in systems neuroscience*, *2*(November), 4. doi:10.3389/neuro.06.004.2008

Kruskal, J. B. (1965). Analysis of Factorial Experiments by Estimating Monotone Transformations of the Data. *Journal of the Royal Statistical Society. Series B (Methodological)*, *27*(2), 251–263. doi:10.2307/2984194

Kuhberger. (1998). The Influence of Framing on Risky Decisions :, *75*(1), 23–55.

Lamm, C., Silani, G., & Singer, T. (2015). Distinct neural networks underlying empathy for pleasant and unpleasant touch. *Cortex*. doi:10.1016/j.cortex.2015.01.021

Lammel, S., Ion, D. I., Roeper, J., & Malenka, R. C. (2011). Projection-Specific Modulation of Dopamine Neuron Synapses by Aversive and Rewarding Stimuli. *Neuron*, *70*(5), 855–862. doi:10.1016/j.neuron.2011.03.025

Lamy, D., Alon, L., Carmel, T., & Shalev, N. (2014). The Role of Conscious Perception in Attentional Capture and Object-File Updating. *Psychological Science*. doi:10.1177/0956797614556777

Lawson, R. P., Seymour, B., Loh, E., Lutti, A., Dolan, R. J., Dayan, P., Weiskopf, N., et al. (2014). The habenula encodes negative motivational value associated with primary punishment in humans. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(32), 11858–63. doi:10.1073/pnas.1323586111

Lebreton, M., Jorge, S., Michel, V., Thirion, B., & Pessiglione, M. (2009). An automatic valuation system in the human brain: evidence from functional neuroimaging. *Neuron*, *64*(3), 431–9. doi:10.1016/j.neuron.2009.09.040

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 : the official journal of the Society for Neuroscience*, *31*(41), 14693–707. doi:10.1523/JNEUROSCI.2218-11.2011

Levy, D. J., & Glimcher, P. W. (2012). The root of all value: a neural common currency for choice. *Current opinion in neurobiology*, *22*(6), 1027–38. doi:10.1016/j.conb.2012.06.001

Li, S., Jiang, X., Yu, H., & Zhou, X. (2014). Cognitive empathy modulates the processing of pragmatic constraints during sentence comprehension. *Social Cognitive and Affective Neuroscience*, *9*, 1166–1174. doi:10.1093/scan/nsu091

Li, Y., Sescousse, G., Amiez, C., & Dreher, J.-C. (2015). Local morphology predicts functional organization of experienced value signals in the human orbitofrontal cortex. *Journal of Neuroscience*, *35*(4), 1648–1658. doi:10.1523/JNEUROSCI.3058-14.2015

Lim, S.-L., O'Doherty, J. P., & Rangel, a. (2013). Stimulus Value Signals in Ventromedial PFC Reflect the Integration of Attribute Value Signals Computed in Fusiform Gyrus and Posterior Superior Temporal Gyrus. *Journal of Neuroscience*, *33*(20), 8729–8741. doi:10.1523/JNEUROSCI.4809-12.2013

Lin, A., Adolphs, R., & Rangel, A. (2012). Social and monetary reward learning engage overlapping neural substrates. *Social cognitive and affective neuroscience*, *7*(3), 274–81. doi:10.1093/scan/nsr006

Lindquist, K. a, Satpute, A. B., Wager, T. D., Weber, J., & Barrett, L. F. (2015). The Brain Basis of Positive and Negative Affect: Evidence from a Meta-Analysis of the Human Neuroimaging Literature. *Cerebral cortex (New York, N.Y. : 1991)*, 1–13. doi:10.1093/cercor/bhv001

Litt, A., Plassmann, H., Shiv, B., & Rangel, A. (2011). Dissociating valuation and saliency signals during decision-making. *Cerebral cortex (New York, N.Y. : 1991)*, *21*(1), 95–102. doi:10.1093/cercor/bhq065

Liu, X., Hairston, J., Schrier, M., & Fan, J. (2011). Common and distinct networks underlying reward valence and processing stages: A meta-analysis of functional neuroimaging studies. *Neuroscience and Biobehavioral Reviews*, *35*(5), 1219–1236. doi:10.1016/j.neubiorev.2010.12.012

Losecaat Vermeer, A. B., Boksem, M. a S., & Sanfey, A. G. (2014). Neural mechanisms underlying context-dependent shifts in risk preferences. *NeuroImage*, *103*, 355–363. doi:10.1016/j.neuroimage.2014.09.054

Luo, S., Ainslie, G., Giragosian, L., & Monterosso, J. R. (2009). Behavioral and neural evidence of incentive bias for immediate rewards relative to preference-matched delayed rewards. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *29*(47), 14820–7. doi:10.1523/JNEUROSCI.4261-09.2009

MacKillop, J., Amlung, M. T., Wier, L. M., David, S. P., Ray, L. a, Bickel, W. K., & Sweet, L. H. (2012). The neuroeconomics of nicotine dependence: a preliminary functional magnetic resonance imaging study of delay discounting of monetary and cigarette rewards in smokers. *Psychiatry research*, *202*(1), 20–9. doi:10.1016/j.pscychresns.2011.10.003

Mackillop, J., Amlung, M. T., Few, L. R., Ray, L. a, Sweet, L. H., & Munafò, M. R. (2011). Delayed reward discounting and addictive behavior: a meta-analysis. *Psychopharmacology*, *216*(3), 305–21. doi:10.1007/s00213-011-2229-0

Manelis, A., & Reder, L. M. (2014). Effective connectivity among the working memory regions during preparation for and during performance of the n-back task. *Frontiers in Human Neuroscience*, *8*(August), 1–11. doi:10.3389/fnhum.2014.00593

Martín-Loeches, M., Hernández-Tamames, J. a, Martín, a, & Urrutia, M. (2014). Beauty and ugliness in the bodies and faces of others: an fMRI study of person esthetic judgement. *Neuroscience*, *277*, 486–97. doi:10.1016/j.neuroscience.2014.07.040

Mattfeld, a. T., Gluck, M. a., & Stark, C. E. L. (2011). Functional specialization within the striatum along both the dorsal/ventral and anterior/posterior axes during associative learning via reward and punishment. *Learning & Memory*, *18*, 703–711. doi:10.1101/lm.022889.111

McClure, S. M., Laibson, D. I., Loewenstein, G., & Cohen, J. D. (2004). Separate neural systems value immediate and delayed monetary rewards. *Science (New York, N.Y.)*, *306*(5695), 503–7. doi:10.1126/science.1100907

McIntosh, a. R. (2000). Towards a network theory of cognition. *Neural Networks*, *13*, 861–870. doi:10.1016/S0893-6080(00)00059-9

McNamee, D., Rangel, A., & O'Doherty, J. P. (2013). Category-dependent and category-independent goal-value codes in human ventromedial prefrontal cortex. *Nature neuroscience*, *16*(4), 479–85. doi:10.1038/nn.3337

Mendez, M. F., Owens, E. M., Jimenez, E. E., Peppers, D., & Licht, E. a. (2013). Changes in personality after mild traumatic brain injury from primary blast vs. blunt forces. *Brain injury : [BI]*, *27*(January), 10–8. doi:10.3109/02699052.2012.722252

Menon, V, & Levitin, D. J. (2005). The rewards of music listening: response and physiological connectivity of the mesolimbic system. *NeuroImage*, *28*(1), 175–84. doi:10.1016/j.neuroimage.2005.05.053

Menon, Vinod, & Uddin, L. Q. (2010). Saliency, switching, attention and control: a network model of insula function. *Brain structure & function*, *214*(5-6), 655–67. doi:10.1007/s00429-010-0262-0

Metereau, E., & Dreher, J.-C. (2014). The medial orbitofrontal cortex encodes a general unsigned value signal during anticipation of both appetitive and aversive events. *Cortex; a journal devoted to the study of the nervous system and behavior*, *63C*, 42–54. doi:10.1016/j.cortex.2014.08.012

Miyapuram, K. P., Tobler, P. N., Gregorios-Pippas, L., & Schultz, W. (2012). BOLD responses in reward regions to hypothetical and imaginary monetary rewards. *NeuroImage*, *59*(2), 1692–9. doi:10.1016/j.neuroimage.2011.09.029

Moscarello, J. M., & LeDoux, J. E. (2013). Active Avoidance Learning Requires Prefrontal Suppression of Amygdala-Mediated Defensive Reactions. *Journal of Neuroscience*, *33*(9), 3815–3823. doi:10.1523/JNEUROSCI.2596-12.2013

Mosher, C. P., & Rudebeck, P. H. (2015). news and views The amygdala accountant : new tricks for an old structure. *Nature Publishing Group*, *18*(3), 324–325. doi:10.1038/nn.3949

Mumford, J. a., & Nichols, T. (2009). Simple group fMRI modeling and inference. *NeuroImage*, *47*(4), 1469–1475. doi:10.1016/j.neuroimage.2009.05.034

Murch, K. B., & Krawczyk, D. C. (2014). A neuroimaging investigation of attribute framing and individual differences. *Social cognitive and affective neuroscience*, *9*(10), 1464–71. doi:10.1093/scan/nst140

Nader, K., Schafe, G. E., & Doux, J. E. L. (2000). Fear memories require protein synthesis in the amygdala for reconsolidation after retrieval. *Nature*, *406*(August).

Neuroscience, H., Boukrina, O., Graves, W. W., Richardson, F. M., & Ruskin, A. (2013). Neural networks underlying contributions from semantics in reading aloud, *7*(September), 1–13. doi:10.3389/fnhum.2013.00518

Niznikiewicz, M. a., & Delgado, M. R. (2011). Two sides of the same coin: Learning via positive and negative reinforcers in the human striatum. *Developmental Cognitive Neuroscience*, *1*(4), 494–505. doi:10.1016/j.dcn.2011.07.006

Northoff, G., Heinzel, A., de Greck, M., Bermpohl, F., Dobrowolny, H., & Panksepp, J. (2006). Self-referential processing in our brain--a meta-analysis of imaging studies on the self. *NeuroImage*, *31*(1), 440–57. doi:10.1016/j.neuroimage.2005.12.002

Odum, A. L. (2011). Delay discounting: I'm a k, you're a k. *Journal of the experimental analysis of behavior*, *96*(3), 427–39. doi:10.1901/jeab.2011.96-423

Ogawa, M., Meer, M. van der, & Esber, G. (2013). Risk-responsive orbitofrontal neurons track acquired salience. *Neuron*, *77*(2), 251–258. doi:10.1016/j.neuron.2012.11.006.Risk-responsive

Orsini, C. a., Trotta, R. T., Bizon, J. L., & Setlow, B. (2015). Dissociable Roles for the Basolateral Amygdala and Orbitofrontal Cortex in Decision-Making under Risk of

Punishment. *Journal of Neuroscience*, *35*(4), 1368–1379. doi:10.1523/JNEUROSCI.3586-14.2015

O'Doherty, J. P. (2011). Contributions of the ventromedial prefrontal cortex to goal-directed action selection. *Annals of the New York Academy of Sciences*, *1239*, 118–29. doi:10.1111/j.1749-6632.2011.06290.x

O'Doherty, J. P. (2014). The problem with value. *Neuroscience and biobehavioral reviews*, 1–10. doi:10.1016/j.neubiorev.2014.03.027

O'Doherty, J. P., Buchanan, T. W., Seymour, B., & Dolan, R. J. (2006). Predictive neural coding of reward preference involves dissociable responses in human ventral midbrain and ventral striatum. *Neuron*, *49*(1), 157–66. doi:10.1016/j.neuron.2005.11.014

O'Doherty, J., Rolls, E. T., Francis, S., Bowtell, R., & McGlone, F. (2001). Representation of pleasant and aversive taste in the human brain. *Journal of neurophysiology*, *85*(3), 1315–21. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11248000

Pereira, C. S., Teixeira, J., Figueiredo, P., Xavier, J., Castro, S. L., & Brattico, E. (2011). Music and emotions in the brain: familiarity matters. *PloS one*, *6*(11), e27241. doi:10.1371/journal.pone.0027241

Peters, E., & Levin, I. P. (2008). Dissecting the risky-choice framing effect : Numeracy as an individual-difference factor in weighting risky and riskless options, *3*(6), 435–448.

Peters, J., & Büchel, C. (2009). Overlapping and distinct neural systems code for subjective value during intertemporal and risky decision making. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *29*(50), 15727–34. doi:10.1523/JNEUROSCI.3489-09.2009

Peters, J., & Büchel, C. (2010). Episodic future thinking reduces reward delay discounting through an enhancement of prefrontal-mediotemporal interactions. *Neuron*, *66*(1), 138–48. doi:10.1016/j.neuron.2010.03.026

Plassmann, H., O'Doherty, J. P., & Rangel, A. (2010). Appetitive and aversive goal values are encoded in the medial orbitofrontal cortex at the time of decision making. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *30*(32), 10799–808. doi:10.1523/JNEUROSCI.0788-10.2010

Qin, J., Kimel, S., Kitayama, S., Wang, X., Yang, X., & Han, S. (2011). How choice modifies preference: neural correlates of choice justification. *NeuroImage*, *55*(1), 240–6. doi:10.1016/j.neuroimage.2010.11.076

Radu, P. T., Yi, R., Bickel, W. K., Gross, J. J., & McClure, S. M. (2011). A mechanism for reducing delay discounting by altering temporal attention. *Journal of the experimental analysis of behavior*, *96*(3), 363–85. doi:10.1901/jeab.2011.96-363

Ramsey, J D, Hanson, S. J., Hanson, C., Halchenko, Y. O., Poldrack, R. a, & Glymour, C. (2010). Six problems for causal inference from fMRI. *NeuroImage*, *49*(2), 1545–58. doi:10.1016/j.neuroimage.2009.08.065

Ramsey, Joseph D, Hanson, S. J., & Glymour, C. (2011). NeuroImage Multi-subject search correctly identifies causal connections and most causal directions in the DCM models of the Smith et al . simulation study ☆. *NeuroImage*, *58*(3), 838–848. doi:10.1016/j.neuroimage.2011.06.068

Rieger, M. O., Wang, M., & Hens, T. (2014). Risk Preferences Around the World Risk Preferences Around the World, (February 2015).

Rothman, a J., & Salovey, P. (1997). Shaping perceptions to motivate healthy behavior: the role of message framing. *Psychological bulletin*, *121*(1), 3–19. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9000890

Roy, M., Shohamy, D., Daw, N., Jepma, M., Wimmer, G. E., & Wager, T. D. (2014). Representation of aversive prediction errors in the human periaqueductal gray. *Nature neuroscience*, *17*(11). doi:10.1038/nn.3832

Roy, M., Shohamy, D., & Wager, T. D. (2012). Ventromedial prefrontal-subcortical systems and the generation of affective meaning. *Trends in cognitive sciences*, *16*(3), 147–56. doi:10.1016/j.tics.2012.01.005

Ruz, M., & Lupiáñez, J. (2002). A review of attentional capture : On its automaticity and sensitivity to endogenous control. *Psicológica*, *23*, 283–309.

Santos, J. P., Seixas, D., Brandão, S., & Moutinho, L. (2011). Investigating the role of the ventromedial prefrontal cortex in the assessment of brands. *Frontiers in neuroscience*, *5*(June), 77. doi:10.3389/fnins.2011.00077

Schaefer, M., & Rotte, M. (2007). Favorite brands as cultural objects modulate reward circuit. *Neuroreport*, *18*(2), 141–5. doi:10.1097/WNR.0b013e328010ac84

Sescousse, G., Caldú, X., Segura, B., & Dreher, J.-C. (2013). Processing of primary and secondary rewards: a quantitative meta-analysis and review of human functional neuroimaging studies. *Neuroscience and biobehavioral reviews*, *37*(4), 681–96. doi:10.1016/j.neubiorev.2013.02.002

Seymour, B., O'Doherty, J. P., Koltzenburg, M., Wiech, K., Frackowiak, R., Friston, K., & Dolan, R. (2005). Opponent appetitive-aversive neural processes underlie predictive learning of pain relief. *Nature neuroscience*, *8*(9), 1234–1240. doi:10.1038/nn1527

Shackman, A. J., Salomons, T. V., Slagter, H. a, Fox, A. S., Winter, J. J., & Davidson, R. J. (2011). The integration of negative affect, pain and cognitive control in the cingulate cortex. *Nature reviews. Neuroscience*, *12*(3), 154–67. doi:10.1038/nrn2994

Sharot, T., Martino, B. D., Dolan, R. J., & De Martino, B. (2009). How choice reveals and shapes expected hedonic outcome. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *29*(12), 3760–5. doi:10.1523/JNEUROSCI.4972-08.2009

Sharot, T., Shiner, T., & Dolan, R. J. (2010). Experience and choice shape expected aversive outcomes. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *30*(27), 9209–15. doi:10.1523/JNEUROSCI.4770-09.2010

Shenhav, A., Botvinick, M. M., & Cohen, J. D. (2013). The expected value of control: an integrative theory of anterior cingulate cortex function. *Neuron*, *79*(2), 217–40. doi:10.1016/j.neuron.2013.07.007

Shenhav, A., & Greene, J. D. (2014). Integrative moral judgment: dissociating the roles of the amygdala and ventromedial prefrontal cortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *34*(13), 4741–9. doi:10.1523/JNEUROSCI.3390-13.2014

Shiner, T., Symmonds, M., Guitart-Masip, M., Fleming, S. M., Friston, K. J., & Dolan, R. J. (2014). Dopamine, Salience, and Response Set Shifting in Prefrontal Cortex. *Cerebral cortex (New York, N.Y. : 1991)*, 1–11. doi:10.1093/cercor/bhu210

Shohamy, D. (2011). Learning and motivation in the human striatum. *Current Opinion in Neurobiology*, *21*(3), 408–414. doi:10.1016/j.conb.2011.05.009

Singer, T., Critchley, H. D., & Preuschoff, K. (2009). A common role of insula in feelings, empathy and uncertainty. *Trends in cognitive sciences*, *13*(8), 334–40. doi:10.1016/j.tics.2009.05.001

Sinha, R., & Li, C. S. R. (2007). Imaging stress- and cue-induced drug and alcohol craving: association with relapse and clinical implications. *Drug and alcohol review*, *26*(1), 25–31. doi:10.1080/09595230601036960

Smith, D. V., Clithero, J. a, Boltuck, S. E., & Huettel, S. a. (2014a). Functional connectivity with ventromedial prefrontal cortex reflects subjective value for social rewards. *Social cognitive and affective neuroscience*. doi:10.1093/scan/nsu005

Smith, D. V., Clithero, J. a, Boltuck, S. E., & Huettel, S. a. (2014b). Functional Connectivity with Ventromedial Prefrontal Cortex Reflects Subjective Value for Social Rewards. *Social cognitive and affective neuroscience*, (919). doi:10.1093/scan/nsu005

Smith, D. V., Hayden, B. Y., Truong, T.-K., Song, A. W., Platt, M. L., & Huettel, S. a. (2010). Distinct value signals in anterior and posterior ventromedial prefrontal cortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *30*(7), 2490–5. doi:10.1523/JNEUROSCI.3319-09.2010

Smith, S. M., Miller, K. L., Salimi-khorshidi, G., Webster, M., Beckmann, C. F., Nichols, T. E., Ramsey, J. D., et al. (2011). NeuroImage Network modelling methods for FMRI. *NeuroImage*, *54*(2), 875–891. doi:10.1016/j.neuroimage.2010.08.063

Strombach, T., Weber, B., Hangebrauk, Z., Kenning, P., Karipidis, I. I., Tobler, P. N., & Kalenscher, T. (2015). Social discounting involves modulation of neural value signals by temporoparietal junction. *Proceedings of the National Academy of Sciences of the United States of America*. doi:10.1073/pnas.1414715112

Sun, J., Hu, X., Huang, X., Liu, Y., Li, K., Li, X., Han, J., et al. (2012). Inferring consistent functional interaction patterns from natural stimulus FMRI data. *NeuroImage*, *61*(4), 987–99. doi:10.1016/j.neuroimage.2012.01.142

Tom, S. M., Fox, C. R., Trepel, C., & Poldrack, R. a. (2007). The neural basis of loss aversion in decision-making under risk. *Science (New York, N.Y.)*, *315*(5811), 515–8. doi:10.1126/science.1134239

Tomasi, D., Volkow, N. D., Wang, G. J., Wang, R., Telang, F., Caparelli, E. C., Wong, C., et al. (2011). Methylphenidate enhances brain activation and deactivation responses to visual attention and working memory tasks in healthy controls. *NeuroImage*, *54*(4), 3101–10. doi:10.1016/j.neuroimage.2010.10.060

Tremblay, L., & Schultz, W. (1999). Relative reward preference in primate orbitofrontal cortex. *Nature*, *398*(6729), 704–8. doi:10.1038/19525

Tversky, A., & Kahneman, D. (1986). Rational choice and the framing of decisions. *Journal of business*, *59*(4). Retrieved from http://www.jstor.org/stable/10.2307/2352759

Tzagarakis, C., Jerde, T. a., Lewis, S. M., Uğurbil, K., & Georgopoulos, A. P. (2009). Cerebral cortical mechanisms of copying geometrical shapes: A multidimensional scaling analysis of fMRI patterns of activation. *Experimental Brain Research*, *194*, 369–380. doi:10.1007/s00221-009-1709-5

Uddin, L. Q. (2014). Salience processing and insular cortical function and dysfunction. *Nature Publishing Group*, *16*(1), 55–61. doi:10.1038/nrn3857

Wang, D. V., & Tsien, J. Z. (2011). Convergent processing of both positive and negative motivational signals by the VTA dopamine neuronal populations. *PLoS ONE*, *6*(2). doi:10.1371/journal.pone.0017047

Webb, J. T., Ferguson, M. a, Nielsen, J. a, & Anderson, J. S. (2013). BOLD Granger causality reflects vascular anatomy. *PloS one*, *8*(12), e84279. doi:10.1371/journal.pone.0084279

Weber, E., Blais, A., & Betz, N. (2002). A domain‐specific risk‐attitude scale: Measuring risk perceptions and risk behaviors. *Journal of Behavioral Decision Making*, *290*(August), 263–290. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/bdm.414/full

Weigard, A., Chein, J., Albert, D., Smith, A., & Steinberg, L. (2014). Effects of anonymous peer observation on adolescents' preference for immediate rewards. *Developmental science*, *17*(1), 71–8. doi:10.1111/desc.12099

Welchew, D. E., Honey, G. D., Sharma, T., Robbins, T. W., & Bullmore, E. T. (2002). Multidimensional scaling of integrated neurocognitive function and schizophrenia as a disconnexion disorder. *NeuroImage*, *17*, 1227–1239. doi:10.1006/nimg.2002.1246

Whalen, P. J., Kagan, J., Cook, R. G., Davis, F. C., Kim, H., Polis, S., McLaren, D. G., et al. (2004). Human amygdala responsivity to masked fearful eye whites. *Science (New York, N.Y.)*, *306*(5704), 2061. doi:10.1126/science.1103617

Willems, R. M., der Haegen, L. V., Fisher, S. E., & Francks, C. (2014). On the other hand: including left-handers in cognitive neuroscience and neurogenetics. *Nature Reviews Neuroscience*, *15*(3), 193–201. doi:10.1038/nrn3679

Wilson-Mendenhall, C. D., Barrett, L. F., & Barsalou, L. W. (2013). Neural Evidence That Human Emotions Share Core Affective Properties. *Psychological Science*, *24*, 947–956. doi:10.1177/0956797612464242

Winecoff, A., Clithero, J. a, Carter, R. M., Bergman, S. R., Wang, L., & Huettel, S. a. (2013). Ventromedial prefrontal cortex encodes emotional value. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *33*(27), 11032–9. doi:10.1523/JNEUROSCI.4317-12.2013

Winston, J. S., Vlaev, I., Seymour, B., Chater, N., & Dolan, R. J. (2014). Relative valuation of pain in human orbitofrontal cortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *34*(44), 14526–35. doi:10.1523/JNEUROSCI.1706-14.2014

Wolff, S. B. E., Gründemann, J., Tovote, P., Krabbe, S., Jacobson, G. a, Müller, C., Herry, C., et al. (2014). Amygdala interneuron subtypes control fear learning through disinhibition. *Nature*, *509*, 453–8. doi:10.1038/nature13258

Wu, C. S., & Cheng, F. F. (2011). The joint effect of framing and anchoring on internet buyers' decision-making. *Electronic Commerce Research and Applications*, *10*(3), 358–368. doi:10.1016/j.elerap.2011.01.002

Yokoyama, R., Nozawa, T., Sugiura, M., Yomogida, Y., Takeuchi, H., Akimoto, Y., Shibuya, S., et al. (2014). The neural bases underlying social risk perception in purchase decisions. *NeuroImage*, *91C*, 120–128. doi:10.1016/j.neuroimage.2014.01.036

Zajonc, R. (1968). Attitudinal effects of mere exposure. *Journal of personality and social psychology*, *9*(2). Retrieved from http://psycnet.apa.org/journals/psp/9/2p2/1/