Neural dynamics of affect, gist, probability, and choice

Action Editor: Angelo Cangelosi

Daniel S. Levine*

Department of Psychology, University of Texas at Arlington, Arlington, TX 76019-0528, United States

Received 29 April 2011; accepted 24 June 2011

Abstract

Recent behavioral data show that the traditional reduction of all probabilistic choices to choices among monetary gambles is inaccurate. Specifically, while decision makers tend to overweight low probabilities of obtaining any resource, the overweighting is greater when the resource is more emotionally evocative. We present a shunting nonlinear neural network that simulates the biasing effect of emotion on probabilistic choice. The network includes analogs of parts of the amygdala, orbitofrontal cortex, ventral striatum, thalamus, and anterior cingulate as well as sensory and premotor cortices. The network classifies prospective probabilistic options by means of an adaptive resonance module with vigilance selective for those attributes that are deemed most significant for the option currently being processed. The categories into which these options are placed embody significant gists of the options in a manner consistent with fuzzy trace theory.

© 2011 Elsevier B.V. All rights reserved.

Keywords: Decision making; Emotion; Amygdala; Orbitofrontal cortex; Adaptive resonance theory; Fuzzy trace theory

1. Introduction

One aspect of human decision making data that deviates from classical economic models is the nonlinear weighting of probabilities. Experiments on choices between explicitly described gambles show that human decision makers tend to overweight low nonzero probabilities and underweight high non-unity probabilities. Tversky and Kahneman’s (1992) prospect theory includes a mathematical formulation of this nonlinear probability weighting, as shown by the S-shaped curve of Fig. 1.

Yet Tversky and Kahneman’s theory tacitly assumed that specific probabilities had the same weights regardless of what resource they dealt with; for example, a certain percentage probability of saving a person’s life, avoiding damage to one’s house, or winning a trip to Europe in a raffle could all be reduced to the same monetary gamble. Against this simplifying assumption, there is significant evidence that the curvature of the S in Fig. 1 is different for different types of resources. For example, Rottenstreich and Hsee (2001) asked some of their participants if they would rather obtain $50 or the kiss of their favorite movie star, and the majority (70%) preferred the money. But when the same participants were given a hypothetical choice between a 1% probability of obtaining the $50 and a 1% probability of obtaining the kiss, the majority (65%) preferred the kiss. Rottenstreich and Hsee explained their finding by noting that the kiss was affect-rich whereas the money was affect-poor. They concluded that a low nonzero probability of obtaining an affect-rich resource is more strongly overweighted than the same low probability of obtaining an affect-poor resource, as described in Fig. 2.

The kiss-money data actually deal with curvature of the S curves only at the end closest to probability 0. Another result by Rottenstreich and Hsee (2001) indicates that affective richness also leads to sharper curvature at the end closest to probability 1. These researchers asked another set of participants how much they would be willing to pay for a 99% probability of obtaining a $500 tuition
rebate (affect-poor) and for a 99% probability of obtaining $500 toward a trip to European tourist destinations (affect-rich: these were American students). The median price that the participants were willing to pay for the almost-certain European trip was $28 lower than the median they were willing to pay for the almost-certain tuition rebate, indicating that the gap between a 99% and a 100% probability was psychologically larger in the affect-rich case.

Variability in the shape of the S-curve is also supported by results of Kusev, van Schaik, Ayton, Dent, and Chater (2009) who framed equivalent probabilistic choices either as insurance purchases – for example, “(a) 1% chance of winning an insurance rebate of £400 or (b) a guaranteed insurance rebate of £20” (Kusev et al., 2009, p. 1489) – all probabilities of gains above .2 but less than 1 were underweighted.

Gonzalez and Wu (1999) fit the S-shaped curve of prospect theory to a specific mathematical function with several cognitively significant parameters. However, these researchers did not include a theory of underlying cognitive or neural processes that generate those parameters. There have been partial mappings of prospect theory to brain processes (Tom, Fox, Trepel, & Poldrack, 2007; Trepel, Fox, & Poldrack, 2005) but these mappings have not yet been integrated into a quantitative model.

Our goal is to develop a neurocognitive theory that can account for characteristic human distortions of probability processing. As part of this theoretical process, we develop and simulate a brain-based neural network model of the Rottenstreich and Hsee (2001) data on probability weighting with affect-rich and affect-poor resources. Our model does not generate an explicit probability weighting curve, but instead treats probabilities as one attribute of complex stimuli that are processed as a whole. The model incorporates elements of several existing theories that have been utilized in the simulation of other cognitive data: the adaptive resonance theory of categorization (Carpenter & Grossberg, 1987); the gated dipole theory of affective contrasts (Grossberg & Gutowski, 1987); and the fuzzy trace theory of memory (Reyna & Brainerd, 2008; Reyna, Lloyd, & Brainerd, 2003). The model also incorporates roles for different prefrontal and limbic regions that are compatible with fMRI results on emotionally influenced decision making (DeMartino, Kumaran, Seymour, & Dolan, 2006).

2. Background and structure of the model

2.1. Fuzzy emotional traces

One of the clues to understanding nonlinear probability weights arises from fuzzy trace theory (Reyna et al., 2003). Fuzzy trace theory posits the coexistence and interaction of two distinct systems for encoding information: literal or verbatim encoding, and intuitive or gist encoding. Verbatim encoding means literal storage of facts, whereas gist encoding means storing the essential intuitive meaning or “gist” of a situation.

As Reyna et al. (2003) note, gist encoding of probabilities tends toward all-or-none representations of risk. That is, the gist encoding perceives gambles as “certainty,” “no chance,” or “some chance” of a particular gain or loss, and the precise probability of that gain or loss is largely neglected. For this reason, gist encoding tends to reduce the relative attractiveness of sure losses and enhance the relative attractiveness of sure gains in comparison with risky alternatives. The S-shaped function of Fig. 1 was interpreted in Levine (2011) as a nonlinear weighted average of an all-or-none step function arising from gist encoding and a linear function arising from gist encoding and a linear function arising.
from verbatim encoding. The difference between affect-poor and affect-rich S-curves (Fig. 2) suggests that the stronger the affective associations of the resource in question, the greater is the relative influence of the gist encoding compared with the verbatim encoding. Although Reyna and Brainard (2008) stress that affect is not intrinsic to gist representations, they also note that gist representations often depend crucially on emotion. The importance of emotions in gist is particularly likely to be strong in preference decisions such as the ones modeled here.

Hence, the gist processing that influences preferences on decision tasks seems likely to engage brain regions involved in emotional encoding; notably, the amygdala and orbitofrontal cortex (OFC). Sander, Grafman, and Zalla (2003) review evidence that a basic function of the amygdala is detection of those stimuli that are relevant for achieving of specific goals. It is a plausible conjecture that this relevance function also extends to deciding which attributes of a stimulus or mental input are relevant, that is, extracting the gists that are deemed to be suitable for the current context. Barbas and Zikopoulos (2006) distinguish functions of the orbital and lateral parts of the prefrontal cortex. They note that orbital prefrontal processes sensory information from relatively late stages of processing whereas lateral prefrontal processes early sensory information. This could be treated as circumstantial evidence that orbital processes gists and lateral recovers some of the verbatim information that was lost at the orbital level.

2.2. Adaptive resonance

The amygdala and OFC are connected by extensive reciprocal pathways (see, e.g., Schoenbaum, Setlow, Saddoris, & Gallagher, 2003). There are more pathways from amygdala to OFC than from OFC to amygdala, and several other models of decision or conditioning data (Dranias, Grossberg, & Bullock, 2008; Frank & Claus, 2006; Litt, Eliasmith, & Thagard, 2008) include only unidirectional amygdala-to-OFC connections. Yet the reverse pathways from OFC to amygdala seem to be important for the consciousness of emotional experience (Barbas & Zikopoulos, 2006). Hence, we assume in our model that OFC-to-amygadal connections play a key role in top-down classification of emotionally significant inputs.

Several results indicate systematic differences between the representations of emotional stimuli at the OFC level and at the amygdalar level. For example, stimulus-reinforcement associations can be more rapidly learned and reversed by OFC neurons than by amygdala neurons (Rolls, 2006). Also, OFC damage but not amygdalar damage abolishes anticipatory autonomic responses to aversive events before they occur (Bechara, Damasio, Damasio, & Anderson, 1994). Both these findings suggest a hierarchical relationship between those two regions, that is, OFC representations of emotional stimuli are more influenced by higher-order cognition than are amygdalar representations of those stimuli.

The hierarchical relationship between amygdala and OFC is somewhat analogous to hierarchical relationships between primary and secondary cortical areas. Hierarchical relationships between levels of sensory processing have been computationally studied and simulated using adaptive resonance theory (ART; Carpenter & Grossberg, 1987; Carpenter, Grossberg, & Reynolds, 1991). For that reason we model the amygdala-OFC connections relevant for emotionally based decision making by means of a modified adaptive resonance module.

The most basic ART network (Fig. 3) consists of two interconnected layers of nodes, called $F_1$ and $F_2$. The nodes at $F_1$ respond to input features, and those at $F_2$ respond to categories of $F_1$ node activity patterns. Synaptic connections between the two layers are modifiable in both directions. $F_2$ nodes compete with one another in a recurrent on-center off-surround network, and the input is tentatively classified in the winning node’s category. Inhibition from $F_2$ to $F_1$ shuts off most $F_1$ activity if the input mismatches the winning category’s prototype. If the match is close, the combination of top-down signals from $F_2$ and input signals bring total $F_1$ activity high enough to inhibit the reset node. If mismatch occurs, by contrast, the reset node is activated and shuts off the winning category node as long as the current input is present, so that other categories can be considered. Matching is said to occur when some function representing correspondence between top-down and bottom-up patterns is greater than a positive constant $r$ called vigilance.

For our decision model, we interpret the level $F_1$ in Fig. 3 as a set of amygadal encodings of attributes of the probabilistic options between which a choice is being made. We interpret $F_2$ as a set of OFC categories of compound events represented by $F_1$ activity patterns. The categorizations at $F_2$ are assumed to represent gists of the
original options, which can be influenced by various contextual elements including affective richness. We interpret the reset node as corresponding to some part of the anterior cingulate cortex (ACC), based on that brain region’s role in conflict or error detection (Botvinick, Braver, Barch, Carter, & Cohen, 2001; Brown & Braver, 2005).

The contexts and environmental stimuli that influence the OFC categorization layer of Fig. 3 could be located in another layer of association cortex that has connections with that categorization layer. In Frank and Claus (2006), such a layer is located in another part of OFC that has stronger connections with other parts of association cortex that represent semantic information. For simplicity, we do not include cortical semantic stores in the current model but telescope these putative interlayer interactions into selective attention weights to the different attributes, weights that are different for each of the two inputs arriving at the amygdalar attribute layer.

2.3. Gated dipoles

How does “reset” take place in an adaptive resonance network? First a tentative choice is made of the category at the upper level $F_2$ in which to classify the input pattern, and that category is the main or the only one activated at that level (Carpenter & Grossberg, 1987). Then if the input is found to sufficiently mismatch the previously stored prototype of that category, activation of that category is disabled. At that time either another category is tried or an uncommitted $F_2$ node is assigned to a new category for the input pattern.

Carpenter and Grossberg (1987) hinted, and Raijmakers and Molenaar (1997) verified mathematically, that deactivation of the mismatched active category node can be accomplished within a real-time network dynamical system if $F_2$ is a dipole field. In a dipole field, each of a set of competing nodes is replaced by a pair of “on” and “off” channels such as in a gated dipole (Grossberg & Gutowski, 1987). The gated dipole network comprises two pathways of opposite cognitive or behavioral significance; for example, one pathway could connote positive affect and the other negative affect, or one pathway could connote the activity of a category representation and the other that representation’s inactivity. Thus it is a network instantiation of the familiar psychological principle of opponent processing (Solomon & Corbit, 1974). The gated dipole utilizes neurotransmitter depletion to make one pathway transiently active (rebound) when there is a decrease of activity.

---

![Diagram of gated dipole network](image)

**Fig. 4.** Generic gated dipole neural network, consisting of “on” (denoted by plus signs) and “off” (denoted by minus signs) pathways of opposite significance, at three successive times. Arrows denote excitation, filled circles inhibition. Darker lines represent greater activity along a pathway. The filled or partially filled squares at the synapses $Z^+$ and $Z^-$ become depleted with use, as indicated by an “anti-Hebbian” term in the differential equations for the weights of those synapses (see the Appendix for equations). These synapses are called gates because their weights represent the fraction of $X_1^+$ or $X_1^-$ activity that is transmitted to the next stage, $X_2^+$ or $X_2^-$. If $X_2^+$.activity is larger than $X_2^-$.activity, $X_3^+$ becomes active (represented by vertical lines inside a square), representing an on response. If $X_2^-$ activity is larger than $X_2^+$.activity, $X_3^-$.becomes active, representing an off response. (a) On response while an input perturbs the on pathway. The gate in the on pathway is more depleted than the gate in the off pathway, but the larger input to the on pathway overcomes the greater depletion, making $X_2^+$.activity larger than $X_2^-$.activity. (b) Rebound off response after the input to the on pathway is shut off. Both pathways are now receiving equal inputs, due to the non-specific arousal, but the on pathway gate is still more depleted than the off pathway gate, making $X_2^-$ activity larger than $X_2^+$.activity. (c) Recovery, in which neither on nor off response occurs. The on pathway gate has recovered from depletion, so both pathways are receiving equal inputs with equal gate weights, making $X_2^+$ and $X_2^-$.activities equal, so that neither $X_3^+$ nor $X_3^-$ is active. (Reciprocal connections from $X_3^+$ to $X_1^-$ are included because Grossberg and Schmajuk (1987) showed they facilitate second-order aversive conditioning.).
in the opposite pathway, as shown in Fig. 4. The rebound property depends on a tonic non-specific arousal signal that perturbs both pathways. In our representation of the Rottenstreich and Hsee (2001) paradigm, the non-specific arousal signal is stronger during presentation of affect-rich inputs than during presentation of affect-poor inputs.

The dipole field at the OFC level could be operating through the medial and lateral parts of the OFC itself, which have been identified as responding to inputs of positive and negative affective valence respectively (Elliott, Agnew, & Deakin, 2008). More likely, this dipole field could be operating through the OFC’s connections with mutually inhibitory neural populations in the ventral striatum. A dipole-like structure arises in the basal ganglia through the direct pathway, which is excitatory in its effects on the cortex, and the indirect pathway, which is inhibitory in its effects on the cortex. In some recent models of decision making tasks (Frank & Claus, 2006; Levine, Mills, & Estrada, 2005), positive emotional inputs activate the direct pathway and acquire positive incentive motivational properties, whereas negative emotional inputs activate the indirect pathway and acquire negative incentive properties. The gated dipole has also been suggested as a framework for modeling the basal ganglia reward system dysfunction underlying Parkinson’s disease, schizophrenia, and other psychiatric disorders (Grossberg, 1984; Mahurin, 1998).

In addition to the OFC dipole field, the network described here includes another dipole field at the amygdalar level that encodes attributes. The attribute dipoles allow the emotional response to an alternative’s level of an emotionally significant attribute to be sensitive to how much that level is larger than the level of the same attribute in competing options. This is because each option being considered resets the steady state of the depletable transmitters, which in turn provides a context for evaluating other options (see Grossberg & Gutowski, 1987, for mathematical details).

2.4. Translating classification to action

Hence our next step is to take the ART network of Fig. 5, with $F_1$ as amygdala and $F_2$ as OFC, and replace
$F_1$ and $F_2$ each by fields of gated dipoles. Yet the resulting network is still incomplete: it classifies potential options in terms of emotionally desirable and undesirable properties, but does not make choices between options. For these classifications to lead to the behavior of choosing one option over another, the ART with dipoles needs to be connected to another network that actually decides on and generates actions. The dipoles in the OFC and amygdala must be linked to areas involved in deciding among actions (the ACC) and in executing those actions (loops involving the striatum, thalamus, and motor areas of the cortex).

Our action network is based on part of Levine et al.’s (2005) network for simulating the Iowa Gambling Task (IGT). Like the IGT network, the network discussed here includes separate modules for response planning and response execution. We place response planning in the ACC, because that region is implicated in attention to action and selection of action (Holroyd, Nieuwenhuis, Mars, & Coles, 2004), and response execution in the pre-motor cortex. The actual options (e.g., probability $p$ of $\$50$ and probability $p$ of kiss) are represented at both the ACC and pre-motor cortex as well as amygdala; in addition, each option has representations at the basal ganglia and thalamus for action gating. The amount of activation sent to the planning and execution nodes corresponding to each option is influenced by the “gist” category into which the OFC-amygdala network classifies that option. That gist is more nuanced for affect-poor than for affect-rich resources, as discussed in Section 3.

Signals carrying the information that particular actions have good or bad consequences pass from the amygdala to the action gate in the ventral striatum. Appetitive signals facilitate the direct pathway; aversive signals facilitate the indirect pathway. The gate opens (a thalamic node fires) when the direct pathway input sufficiently counteracts the indirect pathway input.

3. Network for simulating the Rottenstreich–Hsee data

Fig. 5 shows the complete network for affectively based decision making, which synthesizes ART, gated dipoles, fuzzy traces, and the Iowa Gambling Task model. The network is designed not only to simulate the Rottenstreich–Hsee data but also to be compatible with large-scale cognitive–emotional processes. Modifiable connections from amygdala to OFC enable computation of expected reward or punishment values, in a similar manner to other network models of those regions that reproduce reversal learning and other conditioning data (Dranias et al., 2008; Frank & Claus, 2006). The dipole field at the amygdalar level enables computation of experienced or anticipated emotions to be based partly on comparison of expected outcomes with other possible outcomes, both from the choice made and from competing choices. The dipole field at the OFC level facilitates changes in categorizations of options based on the outcomes of choosing those options.

The interactions between the short-term memory values at the nodes are governed by shunting nonlinear ordinary differential equations. These are a class of equations for interacting neural populations that are analogous to the membrane equations for neurons (Grossberg, 1973). The equations are solved in MATLAB R2006a using fourth-order Runge–Kutta numerical approximation.

In the Rottenstreich–Hsee simulations, there are two inputs to the network representing the options of a probability $p$ of a kiss versus the same probability $p$ of money. It is assumed that each run represents a different experimental participant and that the participant has stored intrinsic positive values of the money and the kiss. (These values do not directly reflect the relative preferences for certain money versus certain kiss; those preferences are discovered by perturbing the network with the kiss and money inputs for $p = 1$.) Intrinsic values of kiss and of money vary randomly from run to run according to normal distributions, with the mean of the distribution being twice as high for the money as for the kiss. The average preference for money was chosen to be larger because typical experimental participants prefer the certainty of money to the certainty of a kiss.

In the simulations, attention shifts back and forth at random times between the two options (inputs) A and B, only one input being attended at a time. The inputs A and B each have three attributes:

- Possible gain (which has a value of 1 or 0);
- Possible non-gain (which also has a value of 1 or 0); and
- Gain probability (which has a continuum of possible values from 0 to 1).

Describe the two options, probability $p$ of kiss and probability $p$ of money, respectively by standard gamble terminology: $(p, \text{kiss})$ and $(p, \text{money})$. Then if $0 < p < 1$, both $(p, \text{kiss})$ and $(p, \text{money})$ are coded by the attribute vector $(1, 1, p)$. If $p = 1$, both are coded by the attribute vector $(1, 0, 1)$. The distinction between the kiss and money options, which allows the two to be coded in different categories by the OFC, occurs in a selective matching criterion to be described below. The matching is based on amygdalar signals to the ACC, signals that have different attribute-selective weights for affect-rich and affect-poor inputs.

Each of the three attributes is represented by a separate dipole at the amygdala. Each amygdalar dipole includes positive and negative input nodes $x^+_i$ and $x^-_i$ (with the index $i$ = 1 for possible gain, 2 for possible non-gain, 3 for probability of gain); depletable transmitters $z^+_i$ and $z^-_i$; next stage nodes $x^+_j$ and $x^-_j$, and output nodes $x^+_k$ and $x^-_k$.

There are five categories at the $F_1$ (OFC) level, each with its own gated dipole. The five categories represent the following classes of prospects:

1. sure gain;
2. sure non-gain;
3. tossup between gain and non-gain;
than 0 and less than 1 are lumped together as abilities, particularly when arousal is high. That is, at the decision makers tend to blur distinctions between probabilities, particularly when arousal is high. Yet the fuzzy trace hypothesis says that, at the gist level, all prospects with probabilities of gain greater than 0 and less than 1 are lumped together as “some chance of gain, some risk of no gain.” The category into which that would place those prospects most naturally is 3, the tossup category. Categories 4 and 5, which involve finer distinctions among probabilities or risks, should not be accessed as readily as categories 1–3, but should become accessible when arousal is low and some deliberation occurs.

Hence the network’s categorization algorithm includes a mechanism for biasing categorizations so that categories 1, 2, and 3 (less nuanced) are favored over categories 4 and 5 (more nuanced). There are two separate features of the network algorithm that favor classification of input options in category 1, 2, or 3. First, competition in the on-center off-surround equations connecting the output nodes $y_{ij}$ for the five categories is biased in favor of categories 1–3 over the other two (see Grossberg & Levine, 1975, for a mathematical theory of biased competition). Second, if the input sufficiently matches one of the prototypes for category 1, 2, or 3, it is placed in that category regardless of whether a better match exists with category 4 or 5.

The matching criteria are differentially weighted by attributes for affect-poor and affect-rich options. Specifically, attribute 3 (precise probability of gain) is unimportant as a match criterion when the kiss is attended but important as a match criterion when money is attended. Let $m_{ki}$ be elements of a weight matrix representing selective attention to different attributes for each input. Let $k$ represent the index of the currently attended input (1 for kiss, 2 for money). For $k = 1$ in the current simulations, because distinctions among probabilities of a kiss are fuzzy and it is the possibility of gain or non-gain that is salient, $m_{11} = m_{12} = 1$ and $m_{13} = .1$.

(1a)

For $k = 2$, the salience is reversed for money; in the simulations shown here, $m_{21} = m_{22} = 4$ and $m_{23} = 1.3$.

(1b)

(Note that the sum across $i$ of the $m_{ki}$ is the same, 2.1, for both $k$ values.)

Fig. 6 shows schematically the differential weighting of attributes in low-emotion and high-emotion cases.

The different attribute selectivities for the kiss and money inputs cause the option (.01, kiss) to be classified in category 3 and the option (.01, money) in category 4. Since the category-4 prototype has a lower value than the category-3 prototype for attribute 3 (probability of gain), those classifications typically make (.01, money) less preferred than (.01, kiss). Conversely, (.99, kiss) and (.99, money) are placed in categories 3 and 5, respectively, making (.99, money) more preferred because the category-5 prototype has a higher value of attribute 3 than the category-3 prototype.

Since there are intrinsic relationships between the attributes and categories, which the participants have presumably learned by the time the study is run, the top-down and bottom-up weights are both initialized at non-zero values. Attributes 1 (possibility of gain) and 3 (probability of gain) are affectively positive and hence should perturb the positive side of the amygdalar dipole, whereas attribute 2 (possibility of no gain) is affectively negative and hence should perturb the negative side of the dipole. The computational algorithm was simplified by making all three attributes perturb the positive side of the dipole but “reverse scoring” attribute 2, so that an input with a certainty of gain had the value 1 in that attribute whereas an input with a possibility of no gain had the small value .1 (not 0 because that would have caused the weights from category 2 to be all zeroes). These considerations led to the following initial (both top-down and bottom-up) weights between each of the categories and the three attributes:

- $j = 1$ (certain gain): $w_{ji} = 1, 1, 1$,
- $j = 2$ (certain non-gain): $w_{ji} = 0, .1, 0$,
- $j = 3$ (tossup between gain and non-gain): $w_{ji} = 1, .1, .5$,
The reset node activity, $R_j$, is influenced by matches between the inputs and top-down weights. For each of the category nodes $j$, the vector of weights from the $j$th node to the 3 attribute nodes is normalized (that is, divided by its Euclidean norm to get a vector of norm 1). The vector of $x_{ki}^j$ node activities is also normalized. Then, depending on which of the two inputs is being attended to at the time ($k = 1$ for kiss, 2 for money), a distance measure is calculated between these two unit vectors that is weighted componentwise by the attribute weights $m_{kj}$ defined above (the exact formula for the distance is given in Appendix A).

Attribute-selective vigilance was utilized by Leven and Levine (1996) in a model of consumer preference behavior, whereby different attributes were selectively emphasized in the “low-arousal” laboratory taste tests more than in the “high-arousal” market. In the current model the attribute weights $m_{kj}$ exert influences at two locations in the network of Fig. 5, both labeled in the figure as “selective attention.” One of those locations is at the signal from the top-down (OFC-to-amygda) prototype weights that branches to the reset and the planning nodes, both assumed to be located in parts of the ACC. The second is in the direct influences of amygdalar attribute representations on the direct and indirect striatal pathways.

This network does not include a specific neural mechanism for generating selective attention weights, but the weights can be broadly explained by interplay between “hot” (high-affect) and “cool” (low-affect) neural systems (e.g., Metcalfe & Mischel, 1999). Note that the weights to attributes 1 and 2 (certain gain and certain non-gain) are higher when the input is affect-rich (high non-specific arousal) and lower when the input is affect-poor (low non-specific arousal). We can posit that the non-specific arousal input to the amygdalar dipoles comes from some hypothalamic or midbrain locus dealing with primary emotions, that is, from part of the hot system. We can also posit that this arousal signal, if intense enough, temporally inhibits the connections between dorsal ACC and dorsolateral prefrontal cortex (DLPFC), which are not shown in Fig. 5. The ACC–DLPFC connections are part of the cool system that becomes active at times of potential rule conflict (Botvinick et al., 2001) and therefore would be expected to be active during decision tasks involving non-trivial concepts such as probabilities. At times when the participant is attending to the money option, arousal is relatively low and the cool cingulate-prefrontal system activates awareness of precise probabilities (attribute 3 in our network), thereby increasing the weight given to that attribute. At times when the participant is attending to the kiss option, arousal is high and the cool system is inhibited, thereby reducing the weight of attribute 3 and increasing the weight of possible gain and possible non-gain (attributes 1 and 2), which are better understood by the hot system. In the discussion we consider future extension of this model to include the DLPFC explicitly.

The same matching criterion is used in the equations for the response planning nodes corresponding to (projected choices of) the two inputs (1% probability of kiss and 1% probability of money. For each of the inputs $k = 1$ or 2, the attribute-weighted distance measure described above between input and top-down weights is calculated for each of the five categories $j = 1, 2, 3, 4, 5$; call it MATCH. There is a vigilance parameter $V_{IG}$, and if any of the MATCH values are below $V_{IG}$, the index of the minimum value is called $JK$ and input $k$ is classified in category $JK$. There is a bias toward the more “accessible” categories 1, 2, and 3, meaning that those categories are tested first to see if their MATCH values are below $V_{IG}$, and the input can be classified in category 4 or 5 only if none of the MATCH values for 1, 2, and 3 are small enough.

Corresponding to each of the two inputs $k = 1$ (kiss) and $k = 2$ (money) are also striatum node activities $S_k$, thalamus node activities $T_k$, and execution node activities $E_k$. There is a sequential set of influences from planning to striatum to thalamus to execution with planning and execution both assumed to be at different cortical layers, corresponding to the familiar basal ganglia-thalamocortical loops (Alexander, DeLong, & Strick, 1986). Also the striatum node is influenced by the amygdala, via signals from the nodes $x_{ki}^j$ weighted by the attribute biases $m_{kj}$.

4. Simulation results

Fig. 7 shows the results of the Rottenstreich–Hsee simulations over 30 runs, each run assumed to represent a participant in the experiment. Mean values over 30 runs were in turn averaged over 5 groups of 30. As shown in that figure, the majority of simulated participants prefer a .01 probability of a kiss to a .01 probability of money, the mean percentage of 63.33% being close to the 65% value for the actual participants in Rottenstreich and Hsee (2001). Also, the mean percentage favoring 100% probability of a kiss to 100% probability of money was 26.67%, which was close to the 30% of the actual participants who preferred the certain kiss to the certain money.

Note also that the percentage of simulated participants favoring a .99 probability of a kiss over a .99 probability of money was about half of the percentage favoring a certain kiss over certain money. This result means, in prospect theory terms, that the probability of .99 was more under-weighted for the kiss than for the money. Greater under-weighting of .99 in the high-affect case was exactly what Rottenstreich and Hsee found when they tested the price participants were willing to pay for a 99% probability of

---

1 As stated in Section 2.2, typical ART networks have a matching function that needs to be greater than a vigilance value for the input to be placed in an existing category. However, any condition that a function be greater than a number can be converted to a condition that another function be less than another number. For these computations it was more succinct to describe the condition in terms of a weighted vector distance being less than a number.
a European trip (high-affect) versus a tuition rebate (low-affect). Hence the network essentially reproduces the Rottenstreich–Hsee data at both ends of the prospect theory probability weighting curve.

Rottenstreich and Hsee (2001) did not run experiments asking participants to choose between probability \( p \) of a kiss and probability \( p \) of \$50 for values of \( p \) intermediate between .01 and .99. Hence, the graph of those probability values in Fig. 7 represents a prediction of what would happen for such choices. The current model predicts that the strong preference for the possibility of a kiss should persist for low probabilities up to about .1, then drop off sharply between .1 and .2. For intermediate probabilities between about .2 and .7, the fraction of participants preferring the possible kiss should be about a third, close to what it is for the certainty of a kiss versus the certainty of money. Then between .7 and .99 that fraction should fall to about a fifth or a quarter favoring the possible kiss.

Simulations were performed with values of the attentional weights \( m_{ki} \) that were different from the values shown in Eqs. (1) but preserved the property of greater weight toward probability with low arousal and toward possibilities with high arousal; that is, \( m_{11} = m_{12} \gg m_{13} \) and \( m_{21} = m_{22} \ll m_{23} \) (\( \gg \) denoting “significantly greater” and \( \ll \) “significantly less.”) For the range of attention weights satisfying those inequalities, the same qualitative shape shown in Fig. 7 was preserved; that is, the preference for the kiss was much greater for probabilities near 0 than for certainty, and slightly less for probabilities near but not equal to 1 than for certainty. Also, the same qualitative shape was preserved when the attentional shifts between the two inputs were made periodic instead of random.

5. Discussion

5.1. Compatibility with cognitive neuroscience data

The network model introduced here was developed to simulate a specific database involving the weighting curve of prospect theory and the influence of affect on the shape of that curve. The fundamental structure of the neural network incorporates several theories that have proved valuable in models of other cognitive phenomena, namely, fuzzy trace theory, adaptive resonance theory, and gated dipole theory.

The network is concurrently designed to capture some of the interactions among brain regions that have been found in various results of decision making experiments in both humans and monkeys. For example, DeMartino et al. (2006) conducted an fMRI study of a human monetary decision task analogous to Tversky and Kahneman’s “Asian Disease Problem,” and compared activation patterns for choices that conformed to the traditional framing effect (risk seeking for losses or risk averse for gains) with choices that violated the framing effect (risk seeking for gains or risk averse for losses). These investigators found that OFC and ACC activity is greater on choices violating the framing effect whereas amygdalar activity is greater on choices conforming to the framing effect.

The network of Fig. 5 can qualitatively explain the DeMartino et al. fMRI data as follows. As in the simulations with the kiss input, the gist representations of prospects involving probabilities of gains and losses tend to be simplified to the four categories of “sure gain,” “risk of gain or no gain,” “sure loss,” or “risk of loss or no loss” (Reyna & Brainerd, 2008). These gist representations favor choices of sure gains over risky gains and choices of risky losses over sure losses, in accordance with the framing heuristic. Consider a network corresponding to a participant who is subject to the framing effect, so presumably his or her decisions are dominated by these simple all-or-none gist representations. Those four gain/loss categories could be encoded at the OFC, and the actual options (e.g., gain \$400 with probability 80%) at the amygdala. The resonant feedback between the amygdala and OFC (see Fig. 3) causes the emotional input to be stably classified in one of those four categories, and the resulting increased \( F_i \) activity inhibits reset (i.e., ACC) activity. Both amygdala
(F\textsubscript{1}) and OFC (F\textsubscript{2}) are activated by the input and the inter-level feedback, but the only F\textsubscript{2} nodes activated are the two corresponding to the “sure gain” and “risky gain” categories; hence, total OFC activity remains relatively small.

By contrast, consider a network corresponding to a second participant not subject to the framing effect, for whom all-or-none gist representations are assumed to be weaker, and either verbatim or more nuanced gist representations stronger, than in the first participant. The network corresponding to that participant will have higher vigilance and may experience mismatches between those inputs and the corresponding simple categories, based on sensitivity to probability and/or magnitude information about the potential gains or losses. That mismatch means that F\textsubscript{1} activity no longer inhibits reset (i.e., ACC) activity. Also, there is greater activity at F\textsubscript{2} (i.e., OFC) than for the first participant because the all-or-none interpretation of the input is challenged and other (existing or novel) categories of gain-loss-probability configurations at that level are considered. The enhanced F\textsubscript{2} activity, in turn, non-specifically inhibits F\textsubscript{1} (i.e., amygdalar) activity, as in Fig. 3. This interpretation of the heuristics is consistent with the suggestion of Trepel et al. (2005) that the prospect theory probability weighting function of Fig. 1 is related to impulsiveness. It is also consistent with the network analysis of Frank and Claus (2006) suggesting that the OFC is required for precise discrimination of magnitudes of reward or punishment.

The network model of Fig. 5 is also compatible with human and monkey choice data showing that the brain executive regions are sensitive to both relative and absolute desirability of options (Blair et al., 2006; Elliott et al., 2008; Tremblay & Schultz, 1999). The gated dipole fields allow for the influence of transmitter depletion from the presentation of other options on the calculated desirability of a particular option. Yet relative preference is not the only influence on the executive system’s response to the options; otherwise, choices between a greater and lesser reward would be treated in the same way as choices between a lesser and a greater punishment. By contrast, the results of Blair et al. (2006) and others indicate that reward/reward choices activate different neural patterns than punishment/punishment choices. These investigators found, for example, that ACC activity was larger in choices between two punishments than in choices between two rewards. The network can capture absolute as well as relative desirability because gated dipoles do not simply “cancel out” positive and negative affective inputs. Rather, dipoles combine these inputs in a nonlinear fashion with dynamics that depend partly on the level of non-specific arousal.

5.2. Other neural models of “prospect theory” data

There are several other neural network models that also establish brain-based accounts of some decision phenomena understood by prospect theory. Among these models are Litt et al. (2008) and Frank and Claus (2006). In general, these models are designed to deal with different dat-
making: in addition to the variability of emotional reactions to gambles with particular probabilities of outcomes, there is variability in the interpretation of the probabilities themselves based on salience of different attributes of each presented alternative. Stone et al. (2003) and others have demonstrated that salience of different aspects of a visual presentation can have a significant effect on risk avoidance. Hence, some types of graphical displays of relative risks lead to different results than numerical presentation of the alternatives. Our analysis would predict that the Rottenstreich–Hsee emotionality effect could similarly be reduced by alternative methods of presentation. For example, if a pie chart were presented and pictures of the kiss or money occupied a wedge of size equal to 1% of the pie, the smallness of the probability of a kiss could become more salient, leading it to be no more attractive than the small probability of money.

The model of Frank and Claus (2006) employs many of the same brain regions as our model to simulate a different database than the one simulated here. Their model reproduces some decision data on a version of the Iowa Gambling Task along with conditioning data such as results on reversal learning and on devaluation. Their model also reproduces framing effects (risk aversion for gains and risk seeking for losses) in a task wherein probabilities are learned from experience.

Frank and Claus, however, recognize that their model is dependent on feedback. Therefore, like the model of Litt et al. (2008), it is not designed to simulate the classical framing effects and other heuristics found by Tversky and Kahneman (1974, 1981) on decisions from description. They note that such effects (which include the Rottenstreich–Hsee data simulated herein) depend on abstract probability concepts and verbal labels, which themselves arise from extensive training. Hence a model such as theirs, or the IGT model of Levine et al. (2005), could be partially synthesized with the current model to form the basis for a larger theory that accounts for decisions both from description and from experience. In particular, the larger model would need to account for how multiatribute gist categories and variable attribute weights arise in development as a result of experience of the world.

5.3. The prospect theory weighting curve

The model described in this paper accounts for the effect of weighting low probabilities more for high-affect stimuli than for low-affect stimuli, yet does not calculate an explicit probability weighting curve such as those shown in Figs. 1 and 2. Rather, probability information is integrated with information about the hypothetical resource to generate emotional responses to entire patterns representing choice options. The responses arise from categorizations of those patterns based on adaptive resonance and fuzzy traces. The model suggests that the probability weighting curve is more a convenient theoretical construct than it is an underlying cognitive or neural reality. Tversky and Kahneman (1981) developed the weighting curve to account for trends in the data, but the curve is not the only possible way to explain those trends. The weighting curve is also absent from the gated dipole decision model of Grossberg and Gutowski (1987), which reproduces many early findings of Tversky and Kahneman (1974, 1981) including preference reversals and framing effects.

Moreover, the probability weighting curve of prospect theory has not yet been explained in terms of underlying cognitive processes that can be fit into a connectionist framework. The authors who perhaps have come closest to that goal are Gonzalez and Wu (1999), who fit the probability weighting curves of prospect theory to the two-parameter class of weighting functions

$$w(p) = \frac{\delta p^c}{\delta p^c + (1 - p)^c}.$$ 

The exponent $c$ in the weighting function $w(p)$ controls curvature and is interpreted as a discriminability parameter. That is, the larger the value of $c$, the greater is the difference between the weights of nearby probabilities over the middle range of probabilities that are not close to either 0 or 1. The multiplier $\delta$ in $w(p)$ controls elevation, that is, how far the curve is above or below the 45° line, and is interpreted as an attractiveness parameter. That is, the larger the value of $\delta$, the more attractive is risk to the decision maker overall. It should be noted that the “attractiveness” interpretation is valid only if the outcome is a potential gain; if the outcome is a potential loss, $\delta$ can be interpreted as risk avoidance.

While Gonzalez and Wu do not provide a mechanistic basis for generating their discriminability and attractiveness parameters, their idea of discriminability bears a striking resemblance to the current model’s selective attention toward the probability attribute, which is larger for low-arousal outcomes. Their attractiveness (or risk avoidance) parameter has no analog in this model but could play a role in simulations of other data to which this model might be extended. For example, the insurance losses in the data of Kusev et al. (2009) exhibit a pattern that corresponds to a large value of $\delta$, whereas the gamble losses in the same data correspond to a small value of $\delta$. In fuzzy trace terms, this pattern might correspond to an unbalanced gist representation; for example, in the insurance context a possible loss might be much more salient than the possible absence of a loss. Such effects might be captured by suitable context-sensitive manipulation of attention weights in the model of Fig. 5.

5.4. Possible applications to other behavioral data

The network of Fig. 5 with some modifications is likely to generalize to a variety of other decision data with two properties: (1) judgments or preferences are based on selective attention or processing of some attributes over others and (2) emotional salience influences attribute selection. Properties (1) and (2) hold for base rate neglect data and ratio bias data, for example.
Base rate neglect (e.g., DeNeys, Vartanian, & Goel, 2008) means that participants make judgments about conditional probabilities on the basis of descriptive information to the exclusion of information about prior probabilities. Effects of salience on base rate neglect were noted by Krynski and Tenenbaum (2007). These researchers verified previous findings that participants provided with the probabilities of women both with and without breast cancer having positive mammograms, and the base rate of breast cancer in the population, then asked to estimate the probability of breast cancer given a positive mammogram, tend to give estimates that are much higher than the value that Bayes’ Rule would indicate. However, Krynski and Tenenbaum showed that if the positive mammogram in non-cancerous women is attributed to a definite source (a benign cyst) rather than to errors or unknown causes, the probability estimate for breast cancer given a positive mammogram becomes close to the Bayes’ Rule value. The authors concluded that non-cancerous positive mammograms became more salient when they had a known cause. The type of network shown in Fig. 5 can capture the influence of salience on base rate neglect or its absence.

Yet these base rate problems differ from Rottenstreich and Hsee (2001) in one important respect: they have objectively correct answers that require reasoning to obtain. In order to model reasoning capabilities, the network of Fig. 5 needs to be expanded to include the dorsolateral prefrontal cortex (DLPFC). The inclusion of attention weights for the probability attribute in our simulations implicitly suggests DLPFC activity (Botvinick et al., 2001) but the model becomes richer when that region is explicitly included.

Brain imaging studies of a base rate task (DeNeys et al., 2008) show that participants who consider base rates exhibit greater DLPFC activity than those who neglect base rates. Cognitive neuroscientists generally believe that when the ACC detects a potential rule conflict or a cause to violate a habitually used rule, it sends a signal to the DLPFC to search short-term memory for information relevant to appropriate rule selection (Botvinick et al., 2001).

Ratio bias means that the same low probability is psychologically perceived as larger when it is the ratio of larger numbers; for example, 10 out of 100 is perceived as larger than 1 out of 10. A consequence of the ratio bias is that many participants who are asked to choose between two options leading to different probabilities of something desirable will choose the smaller probability if it is described by a significantly larger numerator and denominator than those of the other alternative.

For example, Denes-Raj and Epstein (1994) showed their participants two bowls containing red and white jellybeans, told them they would win a certain amount of money if they randomly selected a red jellybean, and instructed them to choose which bowl gave them the best chance of winning money. In one of the bowls, there were a total of 10 jellybeans out of which 1 was red. In the other bowl, there were a total of 100 jellybeans out of which some number greater than 1 but less than 10 were red; hence the probability of drawing red from that bowl was less than 1/10. Like the base rate task, this task has a correct answer that requires reasoning to obtain. A significant number of participants (the majority in the case of choosing 9 out of 100 versus 1 out of 10) were swayed by the larger numbers and chose the bowl with more red jellybeans instead of the one with a higher fraction of red jellybeans.

Levine and Perlovsky (2008) simulated the Denes-Raj and Epstein data using a cognitive model that does not include explicit brain regions but incorporates many of the principles utilized in the network described here. The decision between the two alternative gambles is based on either one of two rules, a heuristic rule based on frequencies and a ratio rule based on probabilities. There is a parameter representing idealized ACC function that determines the likelihood of choosing the ratio rule for a given pair of gambles. If the ratio rule is chosen, another parameter representing idealized DLPFC function determines the probability that the optimal response is made.

In the ratio bias task, the numerator gives a mental indication of the prevalence of the desirable object (e.g., red jellybeans in Denes-Raj & Epstein, 1994), making it more emotionally salient than the denominator. Yet in participants who are somewhat deliberative or have a high need for cognition (Curseeu, 2006), the denominator and numerator are both considered with about equal weight. Even though individual differences play a larger role in the ratio bias task than in the Rottenstreich and Hsee (2001) preference task, we can draw a rough analogy between deliberative use of both numerator and denominator in the ratio bias task and use of explicit probability in the Rottenstreich–Hsee task. This analogy suggest that the ratio bias task could be modeled using a variation of the network of Fig. 5, whereby the idealized ACC and DLPFC parameters of Levine and Perlovsky’s (2008) model are replaced by network variables involving analogs of those areas.

5.5. Summary of modeling principles utilized in the neural network

The following are the main principles employed in the modeling framework of this article, which is intended to apply not just to Rottenstreich and Hsee (2001) but also to a range of other decision data involving emotional and/or deliberative influences on attribute selection.

The amygdala and orbitofrontal cortex (OFC) are connected via a two-layer adaptive resonance network. The amygdalar layer codes emotional values of attributes of chosen response options. The OFC layer codes emotional values of categories or “gists” of options. These categories consist of patterns of attribute values, and the attentional weights attached to attributes are functions of both the task requirement and the decision maker’s personality variables.

The emotional value attached to an option depends partly on current attribute activations multiplied by emotional values of those attributes. Each attribute is encoded
by gated dipoles representing both positive and negative emotional values. The transmitter depletion in the gated dipole networks enables attribute activations of the currently considered option to be compared with attribute activations of imagined alternative options.

The anterior cingulate cortex (ACC) becomes active when gists provide conflicting information about choices between options or rules for such choices. In the cases of networks representing decision makers high in deliberation and/or low in emotional arousal, the in turn activates the dorsolateral prefrontal cortex (DLPFC). DLPFC activity tends to selectively increase the weights of attributes that are relevant to task performance, and decrease the weights of attributes that are task-irrelevant but high in emotional salience.

Appendix A. (Network equations)

During the times when attention is paid to input \(I_k\), where \(I_1 = A\) and \(I_2 = B\), the activities of the input nodes \(x^+_i\) and \(x^-_i\) in the \(i\)th amygdalar attribute dipole, \(i = 1, 2, 3\), satisfy the equations

\[
\frac{dx^+_i}{dt} = -x^+_i + J_i + bx^+_i + (2 - .5k)\chi - R
\]

(2a)

\[
\frac{dx^-_i}{dt} = -x^-_i + bx^-_i + (2 - .5k)\chi - R
\]

(2b)

where \(J_i\) denotes the \(i\)th attribute component of the input vector, \(\chi\) denotes non-specific arousal, with the \(2 - .5k\) factor denoting that arousal is higher (by a factor of 1.5) for the kiss than for the money, and \(R\) denotes activity of the reset node. Eqs. (2a) and (2b) denote that in addition to the node activities are influenced by decay proportional to the activities; excitatory inputs from the dipole output nodes \(x^+_i\) and \(x^-_i\); and inhibition from the reset node. The depletive transmitter weights \(z^+_i\) and \(z^-_i\) satisfy

\[
\frac{dz^+_i}{dt} = (1 - z^+_i) - x^+_i z^+_i
\]

(2c)

\[
\frac{dz^-_i}{dt} = (1 - z^-_i) - x^-_i z^-_i
\]

(2d)

Eqs. (2c) and (2d) denote that the transmitter in the \(i\)th on channel is depleted at a rate proportional to the product of the transmitter weight \(z^+_i\) with the presynaptic node activity \(x^+_i\), and analogously transmitter in the \(i\)th off channel is depleted at a rate proportional to the product of the transmitter weight \(z^-_i\) with the presynaptic node activity \(x^-_i\).

The equations for activities of the nodes at the next layer of the attribute dipole, \(x^+_2\) and \(x^-_2\), are

\[
\frac{dx^+_2}{dt} = -x^+_2 + x^+_1 z^+_1
\]

(2e)

\[
\frac{dx^-_2}{dt} = -x^-_2 + x^-_1 z^-_1
\]

(2f)

denoting that each of those activities decays at a rate proportional to itself and that the node is excited in proportion to the product of presynaptic node activity with transmitter weight. The equations for activities of the nodes at the dipole output layer, \(x^+_3\) and \(x^-_3\), are

\[
\frac{dx^+_3}{dt} = -x^+_3 + (1 - x^+_3) \left( x^+_2 + \sum_{j=1}^{5} y^+_j w^+_i \right) - x^+_3 z^+_3
\]

(2g)

\[
\frac{dx^-_3}{dt} = -x^-_3 + (1 - x^-_3) z^-_3 - x^-_3 z^-_3
\]

(2h)

where \(y^+_j\) denotes activity of the \(j\)th category node; and \(w^+_i\) denotes the weight of the connection between the \(j\)th \(y^+_j\) category node and the \(x^+_3\) node corresponding to the \(i\)th attribute. Eqs. (2g) and (2h) denote that each of those two nodes is excited by the layer-2 dipole node from the same channel and inhibited by the layer-2 node from the opposite channel. In addition, the on-channel output node \(x^+_3\) is excited by inputs from all of the category nodes, \(y^+_j\), weighted by the corresponding connection weights \(w^+_j\). Both (2g) and (2h) are special cases of shunting equations for short-term memory (STM) activity (Grossberg, 1973), which are derived from membrane equations for neurons. The standard shunting equation for the STM activity \(V_i\) of the \(i\)th node in a network is

\[
\frac{dV_i}{dt} = -V_i + (1 - V_i) E_i - V_i I_i
\]

(3)

where \(E_i\) denotes the sum of all excitatory influences on that node and \(I_i\) the sum of all inhibitory influences. If the probability (of either kiss or money) in the currently considered option is \(p\), \(0 < p < 1\), the input vector is taken as \([1, 1, p]\). The first component means that the value of attribute 1 (possibility of gain) is to be 1. The second component is .1 because the value of attribute 2 (possibility of no gain) is also 1, but since this attribute is affectively negative and \(x^+_3\) is an affectively positive site, that attribute is “reverse scored” so that 1 becomes close to 0 (not equal to 0 because that would lead to a category prototype with all zeroes, generating some computational difficulties). The third component is the value of attribute 3 (actual probability of gain). If the probability of kiss or money is 1, then the 0 value of attribute 2 (i.e., no possibility of no gain) is reverse scored as 1, leading to an input vector of \([1, 1, 1]\).

The node activity and transmitter weight equations for the category dipoles satisfy dipole equations analogous to (2a)-(2h), with \(y^+_j\) and \(y^-_j\) being input nodes, \(z^+_i\) and \(z^-_i\) depletive transmitters, \(v^+_j\) and \(v^-_j\) layer-2 nodes, and \(y^+_j\) and \(y^-_j\) output nodes, for \(j = 1-5\), as follows:

\[
\frac{dy^+_j}{dt} = -y^+_j + y^+_j + .5\chi - R
\]

\[
\frac{dy^-_j}{dt} = -y^-_j + y^-_j + .5\chi - R
\]

\[
\frac{dv^+_j}{dt} = \left( 1 - v^+_j \right) - \delta v^+_j
\]

\[
\frac{dv^-_j}{dt} = \left( 1 - v^-_j \right) - \delta v^-_j
\]
The equation for reset node activity is

\[
\frac{dR}{dt} = -R + \min_{j=1}^{5} \text{MATCH}(j),
\]

where MATCH(j) is a measure of closeness between the normalized weight and input vectors, defined as follows: for each \( j = 1, 2, 3, 4, 5 \), and \( i = 1, 2, 3 \),

\[
\text{MATCH}_j = \frac{\sum_{i=1}^{5} w_{ji}}{\sqrt{\sum_{i=1}^{5} (w_{ji})^2}}
\]

\[
\text{NORMINPUT}_i = \frac{I_i}{\sqrt{\sum_{m=1}^{3} (I_m)^2}}.
\]

with \( (I_1, I_2, I_3) \) being the components of the input vector, and

\[
\text{MATCH}(j) = \frac{\sum_{i=1}^{3} m_{ji} (\text{NORMWTS}_j - \text{NORMINPUT}_i)^2.}
\]

The planning node activity corresponding to input \( k \), called \( P_k \), satisfies a shunting equation of the form (3), with a decay term a term representing excitatory influences, and a term representing inhibitory influences:

\[
\frac{dP_k}{dt} = -eP_k + \gamma (1 - P_k) \text{VAL}(k) \sum_{j=1}^{3} w_{jk} m_{ji}
\]

\[
- \phi P_k \text{VAL}(l) \sum_{j=1}^{3} w_{jk} m_{ji}
\]

where \( JK \) is the index of the category into which input \( k \) is classified (see the text for the vigilance algorithm): \( \text{VAL}(k) \) denotes the intrinsic value to the participant of the resource associated with input \( k \) (kiss if \( k = 1 \), money if \( k = 2 \)), \( l \) denotes the index of the other input (so \( l = 3 - k \)); and \( m_{ji} \) are the attribute weights for that input as discussed in the text.

For the simulations described here, the intrinsic value of a kiss, \( \text{VAL}(1) \), is set to \( 10 + 5(\text{randn}) \), and the intrinsic values of money, \( \text{VAL}(2) \), to \( 20 + 10(\text{randn}) \), with \( \text{randn} \) denoting a unit normal random variable that changes value every time it is called. The attribute weights for \( k = 1 \) (kiss) are \( m_{1i} = \{1.1.1\} \), and the attribute weights to \( k = 2 \) (money) are \( m_{2i} = \{4.4.1.3\} \).

If none of the MATCH values is below \( VIG \), so the input is not yet classified, the second and third terms in the differential equation for \( P_k \) is replaced by 0.

The equations for the striatum, thalamus, and execution node activities are shunting equations:

\[
\frac{dS_k}{dt} = -S_k + (1 - S_k) \left( 20P_k + .1 \sum_{i=1}^{3} x_{ji} m_{ki} \right) - S_kSI
\]

\[
\frac{dT_k}{dt} = -T_k + 20(1 - T_k)S_k - T_kTI
\]

\[
\frac{dE_k}{dt} = \frac{1}{\epsilon} (1 - E_k + 10(1 - E_k)T_k - E_kE_l)
\]

where \( l = 3 - k \) as above.

The other parameters listed in the equations are positive constants. For the simulations described here, the values of those parameters were \( b = .2, c = .2, a = 10, \delta = 5, \phi = 50, \epsilon = .05, \gamma = .5, VIG = .05 \). Attention to each input lasted for periods of five time units, and a total run (corresponding to one experimental participant) lasted 50 time units.

References


L. Rauch (Eds.), The orbitofrontal cortex (pp. 57–80). Oxford: Oxford University Press.


