

# From Animal Model to Human Brain Networking: Dynamic Causal Modeling of Motivational Systems

Tal Gonen,<sup>1,2\*</sup> Roee Admon,<sup>1,3\*</sup> Ilana Podlipsky,<sup>1</sup> and Talma Hendler<sup>1,2,3</sup>

<sup>1</sup>Functional Brain Center, Wohl Institute for Advanced Imaging, Tel Aviv Sourasky Medical Center, <sup>2</sup>Department of Psychology, Tel Aviv University, and <sup>3</sup>Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv 64239, Israel

An organism's behavior is sensitive to different reinforcements in the environment. Based on extensive animal literature, the reinforcement sensitivity theory (RST) proposes three separate neurobehavioral systems to account for such context-sensitive behavior, affecting the tendency to react to punishment, reward, or goal-conflict stimuli. The translation of animal findings to complex human behavior, however, is far from obvious. To examine whether the neural networks underlying humans' motivational processes are similar to those proposed by the RST model, we conducted a functional MRI study, in which 24 healthy subjects performed an interactive game that engaged the different motivational systems using distinct time periods (states) of punishment, reward, and conflict. Crucially, we found that the different motivational states elicited activations in brain regions that corresponded exactly to the brain systems underlying RST. Moreover, dynamic causal modeling of each motivational system confirmed that the coupling strengths between the key brain regions of each system were enabled selectively by the appropriate motivational state. These results may shed light on the impairments that underlie psychopathologies associated with dysfunctional motivational processes and provide a translational validity for the RST.

## Introduction

Many animal studies suggest that an organism's behavior is modulated by their individual sensitivity to different reinforcers (for review, see Cardinal et al., 2002). Gray and McNaughton (2000) formulated these animal findings into a single comprehensive and influential model: the reinforcement sensitivity theory (RST), which attempts to account for the observed variability in an organism's behavior. This model includes three major motivational systems mediating behavioral responses to environmental stimuli: (1) the fight-flight-freeze system (FFFS), which mediates sensitivity to aversive stimuli such as punishment, to yield defensive approach (fight) or defensive avoidance (flight, freeze) behaviors; (2) the behavioral activation system (BAS), which mediates sensitivity to appetitive stimuli such as reward to facilitate approach behavior; and (3) the behavioral inhibition system (BIS), which is sensitive to goal-conflict situations such as stimuli of mixed or ambiguous values, and produces adaptive behavioral selection while inhibiting alternative plans.

Findings from animal studies have also identified three separate neural networks underlying the proposed motivational sys-

tems. Specifically, the FFFS, mediating response to aversive stimuli, is thought to involve activation of the periaqueductal gray, medial hypothalamus, central amygdala, and subgenual anterior cingulate cortex (sgACC). The BAS, mediating response to appetitive stimuli, is suggested to involve activation of the ventral tegmental area (VTA), nucleus accumbens (NAcc) and dorsomedial prefrontal cortex (dmPFC). Finally, the BIS is thought to rely on activations in the septohippocampal system and the ventromedial prefrontal cortex (vmPFC).

Although the RST model is appealing as a theoretical framework, and is supported by extensive empirical findings, it was established solely on evidence from animal studies, and thus its generalization to humans is not obvious (Smillie et al., 2007). Recent findings from functional MRI (fMRI) studies suggest that several functions assigned by the RST to specific brain regions are relevant in humans as well, especially the amygdala in response to aversive stimuli and the VTA and NAcc in response to appetitive stimuli (O'Doherty et al., 2002; Phelps, 2006; Haber and Knutson, 2010). Nevertheless, the translation of reinforcement paradigms to human study remains incomplete (Avila et al., 2008).

The aim of the current study was to simultaneously probe the neural networks that underlie human motivational processes as they function in real-life situations and to examine whether they correspond to those proposed by the RST model. To achieve this aim, we used an interactive game that included distinct expressions of relevant motivational states and that has been previously shown to elicit increased amygdala response to punishment and NAcc response to reward (Kahn et al., 2002; Assaf et al., 2009; Admon et al., 2012). We analyzed the individual pattern of neural responses to these stimuli with dynamic causal modeling (DCM) (Friston et al., 2003). This analysis allowed us to estimate the

Received Dec. 13, 2011; revised Jan. 31, 2012; accepted Feb. 7, 2012.

Author contributions: R.A. and T.H. designed research; R.A. performed research; I.P. contributed unpublished reagents/analytic tools; T.G., R.A., and I.P. analyzed data; T.G., R.A., and T.H. wrote the paper.

This work was supported by the Israeli Ministry of Science and Sport (T.H.), the Israeli Defense Forces Medical Corps (T.H.), the Levy Edersheim Gitter Institute for Neuroimaging (T.G. and R.A.), and the Adams Super Center for Brain Studies, Tel Aviv University (R.A. and T.H.).

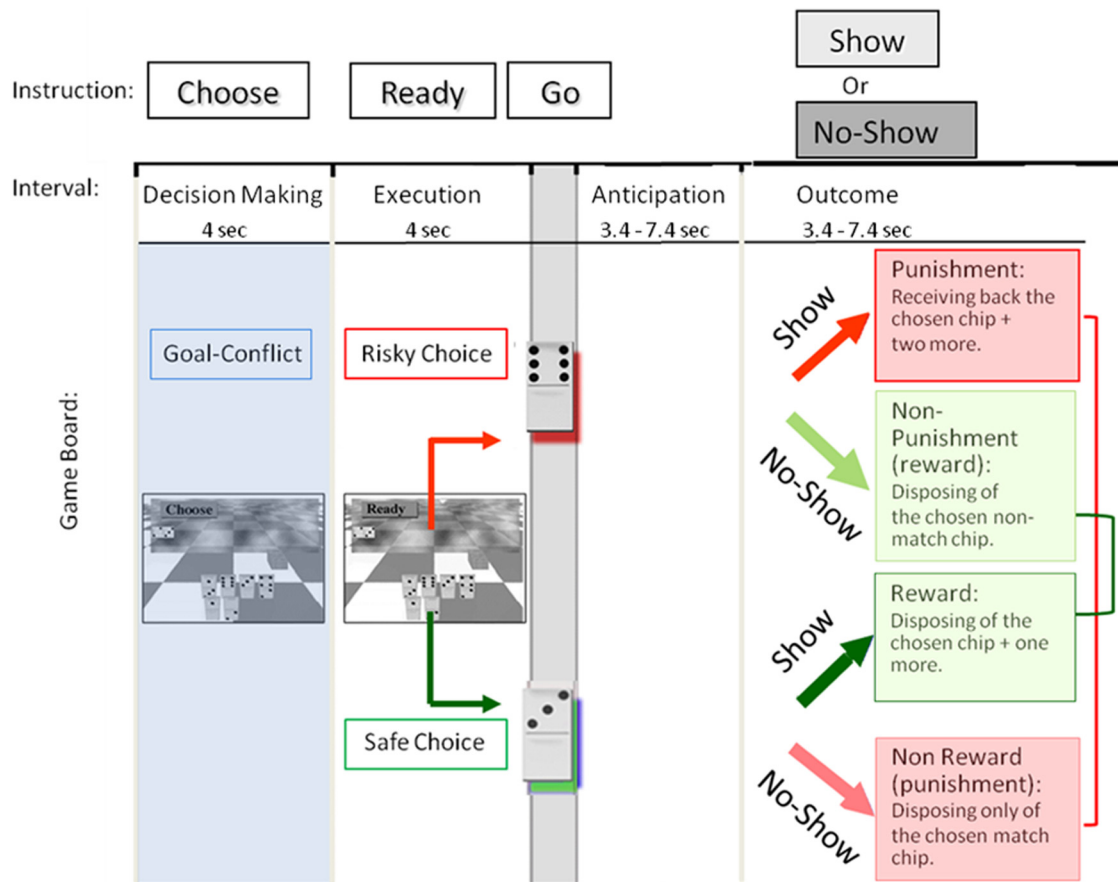
The authors declare no financial conflicts of interests.

\*T.G. and R.A. contributed equally to this work.

Correspondence should be addressed to Talma Hendler, the Functional Brain Center, Wohl Institute for Advanced Imaging, Tel Aviv Sourasky Medical Center, Weizmann 6, Tel-Aviv 64239, Israel. E-mail: talma@tasmc.health.gov.il.

DOI:10.1523/JNEUROSCI.6188-11.2012

Copyright © 2012 the authors 0270-6474/12/327218-07\$15.00/0



**Figure 1.** Domino game paradigm. Each round of the game was composed of four intervals: the player chose which chip to play next (first interval: “Choose”; 4 s), moved the cursor to the chosen chip, and placed it face down adjacent to the master chip (second interval: “Ready” and “Go”; 4 s). The player then waited for the opponent’s response (third interval: “Anticipation”; jittered randomly to 3.4, 5.4, or 7.4 s), and saw whether the opponent challenged this choice by uncovering the chosen chip or not (fourth interval: “Outcome”; jittered randomly to 3.4, 5.4, or 7.4 s). The player’s choices and opponent’s responses were interactively determined by the flow of the game round after round, creating a natural and unpredictable progression of a game situation that lasted 4 min or until the player won. Each player played consecutively for 14 min (average number of games  $\pm$  SEM:  $4.23 \pm 0.06$ ).

strength of coupling within each neuronal network or subgraph corresponding to a motivational system and, crucially, how these connection strengths were modulated by the different motivational states of punishment, reward, and goal-conflict.

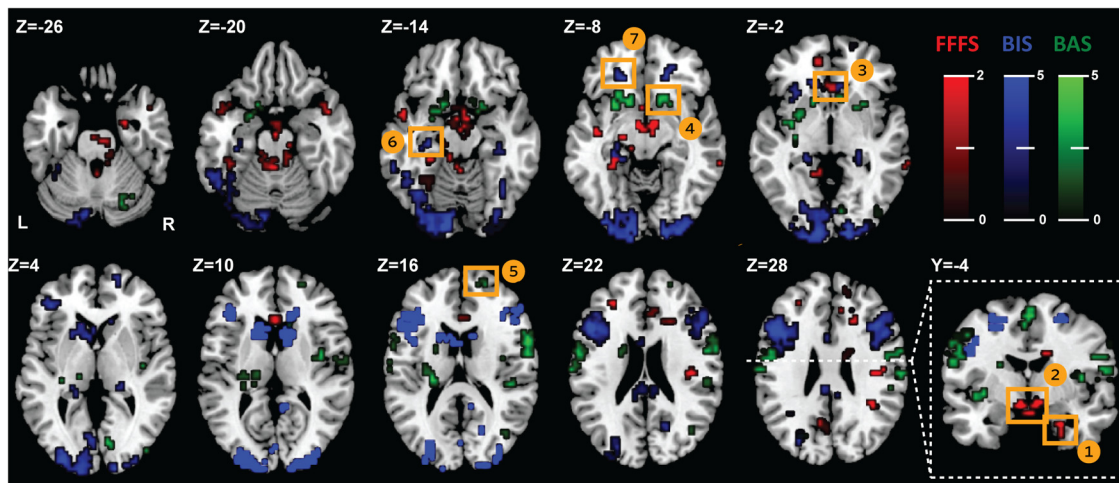
## Materials and Methods

**Participants.** We studied 24 healthy 18-year-old participants (12 males). All were right-handed, had no reported history of psychiatric or neurological disorders, no current use of psychoactive drugs, and no family history of major psychiatric disorders. The protocol was approved by the Tel Aviv Sourasky Medical Center Ethics Committee. All participants provided written informed consent before participation.

**fMRI paradigm.** Participants played a two-player competitive domino game where the opponent’s responses were randomly generated by a computer in a predetermined pattern to allow a balanced design. Players, however, were told that the opponent was the experimenter and that their choices could increase their chances of winning. At the beginning of each game, 12 random domino chips were assigned to the player and were shown on the bottom part of the board, while one master domino chip, which remained constant throughout the game, appeared on the top left corner of the board. Players won the game if they were able to successfully dispose of all 12 chips within 4 min. Each assigned chip could either match the master chip (have one of the master chip’s numbers) or not. In each round of the game, players had to choose one chip (i.e., decision making), place it face down adjacent to the master chip (i.e., execution), and then wait for the opponent’s response (i.e., anticipation) to see whether the opponent challenged this choice by uncovering the chosen chip or not (i.e., outcome). Since the master chip remained con-

stant throughout the game, it was only possible to win by choosing both matching and nonmatching chips. In the game context, matching chips are considered safe moves, since they are associated with rewards if uncovered and nonmatching chips are considered risky moves, since they are associated with punishments if uncovered. Specifically, based on the player’s choice and opponent’s response, there are four possible consequences per game round (i.e., outcome possibilities): (1) show of a nonmatch chip: the choice of a nonmatch chip is exposed and the player is punished by being given the selected chip back, plus two additional chips from the deck; (2) no show of a nonmatch chip: the choice of a nonmatch chip remains unexposed and only the selected chip is disposed of, so the player is not punished; (3) show of match chip: the choice of a match chip is exposed and the player is rewarded by disposal of the selected chip and one additional random chip from the game board; and (4) no show of a match chip: the choice of a match chip is not exposed and only the selected match chip is disposed of, so the player is not rewarded. Overall, player’s choices and opponent’s responses are interactively determined by the flow of the game round after round, creating a natural progression of the game situation that lasts 4 min or until the player wins. Each player played consecutively for 14 min (average number of games  $\pm$  SEM:  $4.23 \pm 0.06$ ). For more details of the game, see Figure 1, as well as Kahn et al. (2002).

**Motivational states.** To probe the neural representation of FFFS, we contrasted the brain’s response to punishment (and non-reward) outcomes to the response to reward (and non-punishment) outcomes in the game (i.e., opponent’s “show” following a player’s nonmatch choice or “no-show” following a match choice vs opponent’s “show” following player’s match choice or “no-show” following a nonmatch choice). To probe the neural representation of BAS, we contrasted the opposite re-



**Figure 2.** Whole-brain analysis. Each of our three contrasts of interest elicited a differential pattern of distributed brain activations that highly corresponded to the motivational neural systems as proposed by the RST model. During response to punishment, activations were observed in the amygdala (1), hypothalamus (2), and sgACC (3), corresponding to the FFFS (red). During response to reward, activations were observed in the NAcc (4) and dmPFC (5), corresponding to BAS (green). Finally, during periods of goal-conflict, activations were observed in the hippocampus (6) and vmPFC (7), corresponding to BIS (blue).  $n = 24$ .

sponses of reward versus punishment (with reward and punishment denoting the same events as in FFFS). The goal-conflict during the “Choose” time interval, in which players were required to decide between the safe, possibly rewarding chip, and the risky, possibly punishing one, was used to probe the neuronal representation entailed by the BIS. Events in which there was no real conflict are rare yet possible (i.e., if only matching chips or only nonmatching chips remained to choose from). When such events occurred, we excluded them by adding a regressor to the general linear model (<2% of our data).

Notably, despite the known theoretical difference between reward and non-punishment, a previous study has shown that in this game, subjects perceive the two states as similarly rewarding; this was also the case for the punishing value of punishment and non-reward (Assaf et al., 2009). To test for differences in the activation level between the different states, we conducted paired  $t$  tests comparing the BOLD percentage signal change (PS) of the two states in each relevant region of interest (ROI; i.e., ROI's defined in the DCM model as receiving input from each state). These  $t$  tests indeed revealed no difference in PS for any of the chosen ROIs in their relevant states: for the amygdala there was no difference in PS between punishment and non-reward states ( $t_{(23)} = 0.17$ ,  $p = 0.86$ ) and for the NAcc there was no difference in PS between reward and non-punishment states ( $t_{(23)} = 0.90$ ,  $p = 0.37$ ). These findings suggest that, at the neural level, the two different states within reward or punishment elicited similar activation during this paradigm.

**fMRI data acquisition.** MRI scans were acquired on a 3.0T MRI scanner (Signa EXCITE; GE Healthcare) with a standard eight-channel head coil using gradient echo-planar imaging sequence of functional T2\*-weighted images (TR, 2500 ms; TE, 35 ms; flip angle, 90°; FOV, 20 × 20 cm; matrix size, 64 × 64) divided into 44 axial slices (thickness, 3 mm with no gap) covering the entire brain.

**fMRI data analysis.** Statistical Parametric Mapping (SPM5; Wellcome Department of Imaging Neuroscience, London, UK) and Marsbar toolbox were used with Matlab 7.6 (MathWorks). Preprocessing of functional scans included slice timing and head-movement correction, normalizing the images to MNI space, and finally spatially smoothing the data (FWHM, 6 mm). In addition, a set of harmonics was used to account for low-frequency noise in the data (1/128 Hz), and the first six images of each functional scan were rejected to allow for T2\* equilibration effects. SPM8 was used with Matlab 7.6.0.324 for the DCM analysis.

**Functional identification of regions of interest.** The size of the effect for each state for each participant was computed using a general linear model (GLM). We included four regressors in the design matrix that corresponded to four stimulus functions convolved with a canonical hemodynamic response function. Three of the GLM regressors encoded our

predetermined motivation estates (i.e., punishment, reward, and conflict), while a fourth model encoded nonspecific effects common to all trials (i.e., common effects). Individual statistical parametric maps were calculated for contrasts of interest testing for the effect of punishment, reward, and conflict. These individual statistical parametric maps were used primarily to identify subject-specific ROI for subsequent effective connectivity analyses using DCM. Specifically, functional time series of the BOLD signal were extracted from each ROI in a subject-specific fashion by placing a sphere of 15 voxels around the individual peak activation of each subject within a group ROI.

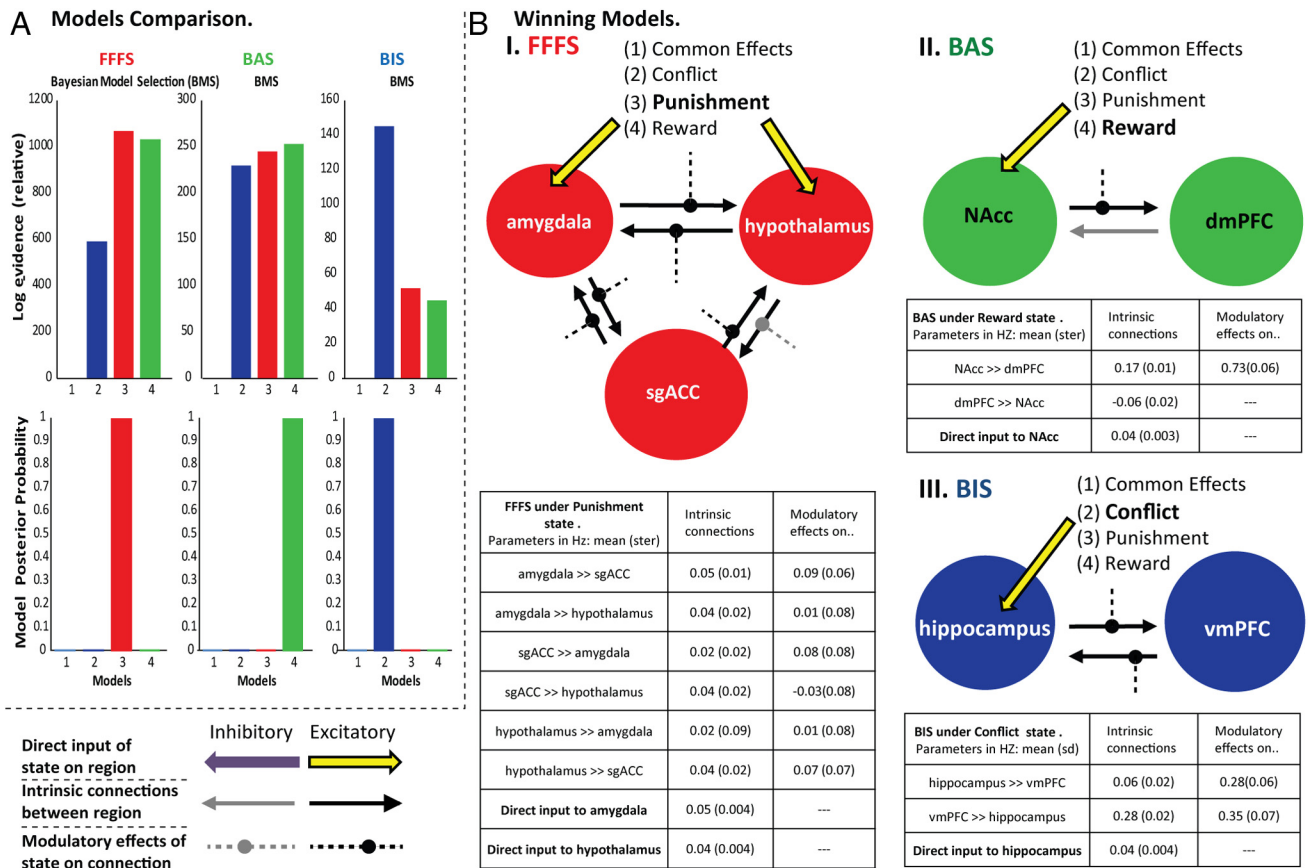
**DCM analysis.** For general information regarding DCM methods, see Friston et al. (2003). In our analyses, we defined four networks (i.e., subgraphs) corresponding to the motivational systems suggested by RST. Within each subgraph, the coupling architecture was the same but differed in terms of which motivational effects changed connection strengths and directly drove regional responses. We identified a significant modulation or enabling of each and every connection by motivational state (punishment, reward, conflict, or common effects) using Bayesian model comparison. Specifically, for the FFFS, we defined bidirectional effective connectivity between the amygdala, hypothalamus, and sgACC with modulatory effects on all connections and direct input to the hypothalamus and amygdala (see Fig. 3B.I). For the BAS, we defined bidirectional effective connectivity between the NAcc and dmPFC, with modulatory effects on the NAcc–dmPFC connection and direct input to NAcc (see Fig. 3B.II). For the BIS, we defined bidirectional effective connectivity between the hippocampus and vmPFC, with modulatory effects on both connections and direct input to the hippocampus (see Fig. 3B.III). Model comparison was performed separately for each motivational subgraph with a fixed effects Bayesian model selection procedure (Penny et al., 2010). Model selection was based on a free energy approximation to model log evidence (Kass and Raftery, 1995), which was used to compute log Bayes factors or log evidence differences. Following Stephan et al. (2010), we considered the evidence for one model over another to be significant when the log evidence exceeded three. This corresponds to a Bayes factor or evidence ratio of  $\exp(3) = 20$ , or a  $p$  value of 0.05. Finally, posterior densities were estimated over models to report the winning model in terms of the one with the highest posterior probability. The parameters of this winning model were averaged over subjects using a Bayesian model averaging procedure (Stephan et al., 2010).

## Results

### Activations of proposed motivational systems

Figure 2 presents an overlay map of activations elicited by our three predetermined independent whole-brain contrasts of inter-





**Figure 3.** DCM results. **A**, Model comparison. Each comparison refers to one subgraph, consisting of four models with the same architecture (i.e., effective and modulatory connections), defined by different motivational states as inputs: (1) common effects, (2) conflict, (3) punishment, and (4) reward. **B**, Winning models. **I**, Effective connections were found to be significant among amygdala, hypothalamus, and sgACC during punishment interval, all suspected to be part of the FFFS. Punishment state had significant input to the amygdala and hypothalamus and modulatory effects on all connections between regions. **II**, Effective connections were found to be significant between NAcc and dmPFC under reward state, as proposed for BAS. Reward state had significant input to NAcc and modulatory effects on the NAcc–dmPFC connection. **III**, Effective connections were significant between the hippocampus and vmPFC during goal-conflict, as proposed for BIS. Goal-conflict state had significant input to the hippocampus and modulatory effects on all connections between regions. Tables display averaged parameter estimates (in Hz).  $n = 24$ .

est. As expected, each contrast elicited a differential pattern of distributed brain activations (Table 1). Importantly for our a priori brain hypothesis, the amygdala, hypothalamus, and sgACC responded only to punishment (Fig. 2, red), while the NAcc and dmPFC responded only to reward (Fig. 2, green), and the hippocampus and vmPFC responded during goal-conflict only (Fig. 2, blue). Notably, although our regions of interest were theoretically predefined by the RST model, these regions also functionally emerged independently from our whole-brain analysis. These results may represent initial evidence that brain responses to each motivational state indeed correspond to the motivational neural systems proposed by the RST model.

**Modeling subgraphs of proposed motivational systems**

The brain regions of each motivational system that are relevant according to the RST model, and that increased their activations in our whole-brain contrasts, were aggregated into three specific subgraphs. Figure 3 shows the winning models according to Bayesian model selection (Fig. 3A). As expected, for each subgraph, the winning model was the one modulated by its assigned RST motivational state. Specifically, for the FFFS, the coupling between the amygdala, hypothalamus, and sgACC was selectively modulated during punishment but not during reward or conflict states. Punishment had a significant input to the amygdala and hypothalamus and also modulated their effective connectivity (Fig. 3B.I). Similarly, for the BAS, the effective connectivity be-

tween the NAcc and dmPFC was selectively modulated while receiving reward, and not while receiving punishment or during goal-conflict. The reward state drove the NAcc and was the only one to modulate these connections (Fig. 3B.II). Finally, for the BIS, the effective connectivity between the hippocampus and vmPFC was selectively modulated when participants were facing a goal-conflict state but not during punishing or rewarding outcomes. Goal-conflict state showed significant input to the hippocampus and modulated the connections between these regions (Fig. 3B.III).

**Discussion**

By manipulating both aversive and appetitive drives, as well as the process of choosing between them under goal-conflict, we were able to probe the neural substrate of distinct motivational processes in humans. Further, simultaneous manipulation of all these states in an unpredictable manner better simulated a real-life situation. Finally, by using a model-driven analysis, we were able to test in humans a well established neural theory of motivation whose construction was based on extensive animal findings. Together, our results display a remarkable correspondence to the proposed RST motivational networks, while highlighting specific regions as core elements in the operation of the FFFS, BAS, and BIS. Furthermore, DCM analyses confirmed that connectivity strength between neural components of each motivational system increased under the relevant state. As defined by Friston

**Table 1. Peak of activations elicited by our three predetermined independent whole-brain contrasts of interest (i.e., motivational states)**

Region	Cluster (no. of voxels)	Peak voxel (x, y, z)	Z value (df = 23)	p < (uncorrected)
<b>Response to punishment outcome</b>				
R hypothalamus	22	6, -6, -12	3.54	0.000
L hypothalamus	20	-6, -6, -9	3.49	0.000
Midbrain	18	0, -13, -21	2.36	0.009
R insula	12	41, -19, 24	3.19	0.001
R amygdala	6	25, -3, -27	3.18	0.001
R inferior temporal gyrus	10	47, 3, -42	2.86	0.002
L fusiform gyrus	10	-28, -53, -12	2.83	0.002
R anterior cingulate cortex (BA 24)	12	0, 25, 9	2.74	0.003
R temporal pole (STG)	7	44, 9, -21	2.72	0.003
L precuneus	10	6, -63, 30	2.62	0.004
R posterior cingulate cortex	7	13, -6, 30	2.52	0.006
L parahippocampal gyrus	11	-28, -34, -12	2.51	0.006
L superior temporal gyrus	6	-47, -16, -6	2.47	0.007
R pons	6	6, -19, -27	2.46	0.007
L cerebellum	21	-6, -41, -18	2.44	0.007
R cerebellum	9	13, -31, -21	2.37	0.009
L temporal pole (STG, BA 38)	12	-50, 6, -15	2.13	0.017
R sgACC (BA25)	8	3, 26, -4	2.06	0.02
L anterior cingulate cortex (BA 32)	6	-9, 47, -3	1.96	0.025
<b>Response to reward outcome</b>				
R inferior frontal gyrus (BA 44)	55	59, 9, 18	4.1	0.000
R supra marginal gyrus	16	63, -22, 30	3.91	0.000
L inferior parietal lobule (BA40)	6	-53, -53, 48	3.89	0.000
L precentral gyrus (BA 6)	28	-63, -13, 36	3.87	0.000
L putamen	6	-31, -16, -3	3.79	0.000
L SMA	28	0, 0, 66	3.69	0.000
L NAcc	61	-16, 16, -3	3.64	0.000
R lingual gyrus	8	9, -78, 0	3.58	0.000
R insula	11	41, -3, 9	3.52	0.000
L insula	8	-34, -3, -3	3.50	0.000
R cerebellum	11	28, -66, -30	3.49	0.000
L thalamus	9	-16, -19, 9	3.49	0.000
R precuneus	7	9, -59, 54	3.44	0.000
R caudate nucleus	9	16, 19, 6	3.39	0.000
L caudate nucleus	6	-22, 0, 21	3.36	0.000
R NAcc	28	9, 16, -3	3.35	0.000
R middle occipital gyrus	6	41, -81, 0	3.27	0.001
R precentral gyrus	7	74, -13, 51	3.23	0.001
L precuneus (BA 7)	14	-19, -56, 39	3.23	0.001
L middle frontal gyrus (BA 8)	9	-34, 22, 36	3.23	0.001
R dorsomedial PFC (BA 10)	8	22, 56, 15	3.13	0.001
<b>Response to goal-conflict</b>				
L inferior frontal gyrus	350	-47, 19, 24	5.09	0.000
L hippocampus	40	-22, -34, -6	4.79	0.000
L occipital inferior and middle gyri (BA 17, 18)	707	-16, -91, -15	4.79	0.000
L occipital inferior gyrus (BA 17)	180	13, -97, -3	4.73	0.000
R inferior frontal gyrus	125	41, 22, 27	4.34	0.000
R ventromedial PFC (BA 11)	7	19, 38, -9	3.95	0.000
L ventromedial PFC (BA 11)	40	19, 38, -9	3.93	0.000
L SMA (BA 8)	36	-13, 16, 48	3.8	0.000
R fusiform gyrus	5	38, -63, -15	3.78	0.000
R middle frontal gyrus	5	31, 16, 57	3.74	0.000
R caudate nucleus	29	13, 16, 12	3.62	0.000
L caudate nucleus	64	-13, 22, 6	3.57	0.000
L posterior cingulate cortex (BA 23)	27	-3, -34, 24	3.53	0.000
L precuneus (BA 7)	9	-9, -72, -51	3.42	0.000
R precuneus (BA 7)	9	19, -66, 24	3.34	0.000
R superior frontal gyrus (BA 10)	9	19, 64, 3	3.63	0.001
L anterior cingulate cortex (BA 33)	5	-3, 13, 27	3.15	0.001
R hippocampus	4	38, -28, -15	3.04	0.001
R posterior cingulate cortex (BA 23)	6	6, -34, 24	3.03	0.001
R superior parietal lobule	7	28, -63, 57	2.96	0.002
R SMA	5	13, 19, 45	2.86	0.002

Localization is based on Montreal Neurological Institute (MNI) criteria. Estimated level of activation is described by Z score and P values. Minimal  $p = 0.05$ , uncorrected, random effect, minimum cluster size 135 mm<sup>3</sup> (i.e., 5 voxels). L, Left; R, right; BA, Brodmann area; SMA, supplementary motor area; STG, stomatogastric ganglion.  $n = 24$ .

(2011), the effective connectivity analysis performed by DCM allows inferences to be made regarding the influence that one neuronal component exerts on another in the network under a specific experimental state. Our findings thus imply that the neural operation of each motivational system in the appropriate context is associated with increased effective connectivity between separate components, a mechanism that may allow the coexistence of different motivational systems within the brain. This is somewhat similar to other suggested neural networks that were shown to increase functional coupling when their function was called on, for example, in the case of the hippocampus–vmPFC circuit in relation to fear extinction (Milad et al., 2007), as well as in the difference in neural coupling during emotion regulation, where downregulation versus upregulation of negative feelings recruited different neural networks (Ochsner et al., 2004).

Few studies have examined the neural networks that underlie motivational processes in humans. For example, Camara et al. (2008) investigated functional connectivity during gains and losses in a gambling task, using the NAcc as a seed region for correlation detection. They found similar networks for both states, including the orbitofrontal cortex, amygdala, insular cortex, and the hippocampus. However, such connectivity analyses could not account for a temporal or causal relationship between regions. DCM analyses have been previously used for investigations of complex processes such as retrieval of emotional memories (Smith et al., 2006) or the process of affective prosody (Ethofer et al., 2006). More relevant to our case, Alexander and Brown (2010) applied DCM analysis to investigate the role of the anterior cingulate cortex (ACC) in signaling perception of risk and predicted reward. They found that computation of reward magnitude and error likelihood is independent (i.e., one is not modulated by the other) and that this computation is intrinsic to the ACC and not received from elsewhere. The authors attribute these findings to the role of the ACC in evaluation of risk versus rewards, a function that the RST relates to BIS. Nonetheless, since their DCM models did not include any of the RST-proposed BIS components, and given the RST prediction that the BIS receives information from the ACC during goal conflict, these results may represent one operational node of the BIS network. Notably, this highlights the importance of comprehensive model hypothesis when applying DCM analysis.

To the best of our knowledge, our study is the first comprehensive investigation of the neural networks that underlie motivational processes in humans using sophisticated causal modeling. It should be noted that our whole-brain analysis results were obtained at a relatively liberal (uncorrected) statistical threshold and thus further investigation of motivational networks in humans is still needed to validate our findings. However, given that our study was driven by a model with a clear a priori hypothesis, the whole-brain analyses performed here were not crucial to the identification of the critical nodes of each motivational network but rather aimed to explore other regions involved in the networks. Indeed, some regions within the distributed activations that were elicited in our whole brain analyses are not mentioned in the RST (e.g., the insula and dorsolateral prefrontal cortex in response to reward and goal-conflict states, respectively). On the contrary, some brain regions that were assigned by the RST to the motivational networks were not revealed in our whole-brain analyses (e.g., the VTA and septal area in response to reward and goal-conflict states, respectively). This signifies that the RST neural model is a simplification of the complex response to motivational signals, especially in humans, which probably involves multiple interacting brain regions.

Thus, the subgraphs analyzed in this paper are clearly part of larger networks. This, however, does not represent a problem for our interpretation in the sense that effective connectivity in dynamic causal modeling can be polysynaptic; in other words, it can be mediated vicariously by regions not included in the model. Finally, it should also be noted that individual differences in the sensitivity to different reinforcements constitute an integral part of the RST that is not addressed in this study. Future investigations should include appropriate psychometric measurements of motivational traits to validate the suggested association between neural activation and RST motivational systems at the individual level.

Despite certain limitations, the remarkable resemblance between our findings in humans and the animal-based RST model strengthens the validity of psychological findings in animal research and the feasibility of its translation to human processes. Furthermore, considering the fact that dysfunctional motivational processes have been implicated in several psychopathologies such as anxiety (McNaughton and Corr, 2008), mood (Depue and Iacono, 1989; Johnson, 2005) and personality disorders (Pastor et al., 2007; Völlm et al., 2007), further exploration of animal models of motivational processes and their parallels in the human brain may promote our ability to understand the neural impairments underlying such pathologies.

## References

- Admon R, Lubin G, Rosenblatt J, Stern O, Kahn I, Assaf M, Hendler T (2012) Imbalanced neural responsivity to risk and reward indicates stress vulnerability in humans. *Cereb Cortex*. Advance online publication. Retrieved January 30, 2012. doi: 10.1093/cercor/bhr369.
- Alexander WH, Brown JW (2010) Competition between learned reward and error outcome predictions in anterior cingulate cortex. *Neuroimage* 49:3210–3218.
- Assaf M, Kahn I, Pearlson GD, Johnson MR, Yeshurun Y, Calhoun VD, Hendler T (2009) Brain activity dissociates mentalization from motivation during an interpersonal competitive game. *Brain Imaging Behav* 3:24–37.
- Avila C, Parcet MA, Barrós-Loscertales A (2008) A cognitive neuroscience approach to individual differences in sensitivity to reward. *Neurotox Res* 14:191–203.
- Camara E, Rodriguez-Fornells A, Münte TF (2008) Functional connectivity of reward processing in the brain. *Front Hum Neurosci* 2:19.
- Cardinal RN, Parkinson JA, Hall J, Everitt BJ (2002) Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex. *Neurosci Biobehav Rev* 26:321–352.
- Depue RA, Iacono WG (1989) Neurobehavioral aspects of affective disorders. *Annu Rev Psychol* 40:457–492.
- Ethofer T, Anders S, Erb M, Herbert C, Wiethoff S, Kissler J, Grodd W, Wildgruber D (2006) Cerebral pathways in processing of affective prosody: a dynamic causal modeling study. *Neuroimage* 30:580–587.
- Friston KJ (2011) Functional and effective connectivity: a review. *Brain Connect* 1:13–36.
- Friston KJ, Harrison L, Penny W (2003) Dynamic causal modelling. *Neuroimage* 19:1273–1302.
- Gray JA, McNaughton N (2000) The neuropsychology of anxiety: an enquiry in to the functions of the septo-hippocampal system. Oxford: Oxford UP.
- Haber SN, Knutson B (2010) The reward circuit: linking primate anatomy and human imaging. *Neuropsychopharmacology* 35:4–26.
- Johnson SL (2005) Mania and dysregulation in goal pursuit: a review. *Clin Psychol Rev* 25:241–262.
- Kahn I, Yeshurun Y, Rotshtein P, Fried I, Ben-Bashat D, Hendler T (2002) The role of the amygdala in signaling prospective outcome of choice. *Neuron* 33:983–994.
- Kass RE, Raftery AE (1995) Bayes factors. *J Am Stat Assoc* 90:773–795.
- McNaughton N, Corr PJ (2008) The neuropsychology of fear and anxiety: a foundation for reinforcement sensitivity theory. In: *The reinforcement sensitivity theory of personality* (Corr PJ, ed), pp 44–94. Cambridge: Cambridge UP.

- Milad MR, Wright CI, Orr SP, Pitman RK, Quirk GJ, Rauch SL (2007) Recall of fear extinction in humans activates the ventromedial prefrontal cortex and hippocampus in concert. *Biological Psychiatry* 62:446–454.
- Ochsner KN, Ray RD, Cooper JC, Robertson ER, Chopra S, Gabrieli JD, Gross JJ (2004) For better or for worse: neural systems supporting the cognitive down- and up-regulation of negative emotion. *Neuroimage* 23:483–499.
- O'Doherty JP, Deichmann R, Critchley HD, Dolan RJ (2002) Neural responses during anticipation of a primary taste reward. *Neuron* 33:815–826.
- Pastor MC, Ross SR, Segarra P, Monta s S, Poy R, Molt J (2007) Behavioral inhibition and activation dimensions: relationship to MMPI-2 indices of personality disorder. *Person Individ Differ* 42:235–245.
- Penny WD, Stephan KE, Daunizeau J, Rosa MJ, Friston KJ, Schofield TM, Leff AP (2010) Comparing families of dynamic causal models. *PLoS Comput Biol* 6:e1000709.
- Phelps EA (2006) Emotion and cognition: insights from studies of the human amygdala. *Annu Rev Psychol* 57:27–53.
- Smillie LD, Dalgleish LI, Jackson CJ (2007) Distinguishing between learning and motivation in behavioral tests of the reinforcement sensitivity theory of personality. *Personality and Social Psychology Bulletin* 33:476–489.
- Smith AP, Stephan KE, Rugg MD, Dolan RJ (2006) Task and content modulate amygdala-hippocampal connectivity in emotional retrieval. *Neuron* 49:631–638.
- Stephan KE, Penny WD, Moran RJ, den Ouden HE, Daunizeau J, Friston KJ (2010) Ten simple rules for dynamic causal modeling. *Neuroimage* 49:3099–3109.
- Völlm B, Richardson P, McKie S, Elliott R, Dolan M, Deakin B (2007) Neuronal correlates of reward and loss in cluster B personality disorders: a functional magnetic resonance imaging study. *Psychiatry Research: Neuroimaging* 156:151–167.