

Planning in view of future needs: a Bayesian model of anticipated motivation

Giovanni Pezzulo (giovanni.pezzulo@istc.cnr.it)

ILC-CNR, Via Giuseppe Moruzzi, 1, 56124 Pisa, Italy
ISTC-CNR, Via S. Martino della Battaglia, 44, 00185 Roma, Italy

Francesco Rigoli (francesco.rigoli@istc.cnr.it)

University of Siena, Piazza S. Francesco 7, 53100 Siena, Italy
ISTC-CNR, Via S. Martino della Battaglia, 44, 00185 Roma, Italy

Abstract

Traditional neuroeconomic theories of decision-making assume that utilities are based on intrinsic values of outcomes and that those values depend on how salient are outcomes in relation to the current motivational state. The fact that humans, and possibly also other animals, are able to plan in view of future motivations is not accounted by this view. So far, it is not clear which are the structures and the computational mechanisms employed by the brain during these processes. In this article, we present a Bayesian computational model that describes how the brain considers future motivations and assigns value to outcomes in relation to this information. We compare our model of anticipated motivation with a model that implements the standard perspective in decision-making and assigns value only based on the animal's current motivations. The results of our simulations indicate an advantage of the model of anticipated motivation in volatile environments. Finally we connect our computational proposal to animal and human studies on prospection and foresight abilities and to neurophysiological investigations on their neural underpinnings.

Keywords: prospection, foresight, goal-directed decision-making, model-based, expected utility, anticipatory motivation

Introduction

In line with *expected utility theory* (EUT), most economic and neuroeconomic models view decision-making as aimed at the maximization of expected utility (von Neumann and Morgenstern, 1944). Many studies have investigated the mechanisms through which the brain encodes utility values associated with planning. Recent neuroeconomic and computational models argue that, in goal-directed behavior, motivation might be the substrate according to which specific utilities are assigned to outcomes. It follows that different motivational states may correspond to different utility functions. With this regard, Niv et al. (2006) define motivation as the mapping between outcomes and their utilities, and refer to "motivational states" (e.g. hunger or thirst) as indices of such different mappings, as one in which foods are mapped to high utilities, and another in which liquids have high utilities.

Although this view can explain most cases of utility assignment in goal-directed behavior, it cannot account for some subtle aspects of human (and possibly also non-human) planning, which derive from *prospection* abilities (Buckner and Carroll, 2007; Gilbert and Wilson, 2007; Pezzulo and Castelfranchi, 2009) and the anticipation of future cognitive, motivational and emotional processes. Indeed, during decision-making, not only humans consider how they currently feel or what they think, but (to some extent) they can also anticipate how they will feel or what they will think successively, when

the outcome is delivered (but see Gilbert and Wilson, 2007 for a discussion of sources of errors in those predictions).

In this paper we focus on the ability to anticipate one's own future motivational states. As recognized by Suddendorf and Corballis (1997) in their "mental time travel hypothesis", a critical feature of planning is the subjects' ability to take action in the light of future motivational needs, independently from their current motivational state. For instance, we go to supermarket even when not hungry, since we anticipate that we will be hungry successively. In terms of EUT, the utility of each future outcome may depend on how much we expect to be motivated when we reach it. The evolutionary advantages of anticipating motivations could be related to better adaptivity in complex and dynamic environments.

In this article, we propose a computational theory of how the goal-directed system assigns utility on the basis of anticipation of future motivations. Indeed, in utility assignment these sources of information seem to interact with the intrinsic values of outcomes, contrary to EUT. Furthermore, we touch the issue of which brain structures could implement these mechanisms in human and non-human animals. Our model extends the Bayesian model of goal-directed decision-making proposed by Botvinick and collaborators (Botvinick and An, 2009; Botvinick et al., sub) (*the baseline model* from now on). Like many RL models, the baseline model assigns utility to outcomes based only on current motivation. By adding to the baseline model a component for modeling motivational dynamics (*a motivational forward model*), we make it able to consider its future motivational states during decision-making. Performance of the two models is compared in three illustrative tasks in which consideration of future motivational states is crucial to maximize rewards.

The baseline model

Recently, Botvinick and collaborators (Botvinick and An, 2009; Botvinick et al., sub) proposed a Bayesian model of goal-directed decision-making to represent the goal-directed computations involved in solving Markov Decision Problems; see fig. 1. The model, which we use as our baseline, takes the form of a directed graphical model (Murphy, 2002). Each node represents a discrete random variable (see fig. 1) and each arrow represents the conditional dependence between two random variables. *State* (s) variables represent the set of world state; *action* (a) variables represent the set of available actions; *policy* (π) variables represent the set of ac-

tions associated with a specific state. Finally, *utility* (u) variables represent the utility function corresponding to a given state. Rather than viewing utility as a continuous variable, the baseline model adopts an approach introduced by Cooper (1988) in which utility is represented through the probability of a binary variable. The following linear transformation maps from scalar reward values to $p(u/s_i)$

$$p(u/s_i) = \frac{1}{2} \left(\frac{R(s_i)}{r_{max}} + 1 \right), r_{max} = \max_j |R(s_j)| \quad (1)$$

In situations involving sequential actions, this model uses a technique proposed by Shachter and Peot (1992) which allows to integrate all rewards in a single representation. This is achieved by introducing a *global utility* (u_G) variable:

$$p(u_G) = \frac{1}{N} \sum_i p(u_i) \quad (2)$$

where N is the number of u nodes. Within this model, the utility of alternative courses of action (e.g. a navigation episode in a labyrinth with different rewards in its branches) can be calculated and maximized by a form of probabilistic inference called *reward query*. In short, the probabilistic model first assigns a desired value (i.e. a maximum value of one) to the aggregated utility node u_G . Then, it uses a standard probabilistic inference algorithm (belief propagation, Pearl, 2000) to compute the posterior probabilities of the policy π nodes. Afterwards, the prior probabilities of the policy nodes are replaced by the obtained posterior probabilities, and the inference algorithm is repeated for several trials. The result is that the optimal policy is computed (see Botvinick and An (2009); Botvinick et al. (sub) for more methodological details). For instance, in a double T-Maze, which has the highest reward in its upper right corner, the selected policy will encode “go right twice”.

The baseline model replicates data from many animal experiments, including devaluation (Balleine and Dickinson, 1998), labyrinth navigation, latent learning and detour behavior (Tolman, 1948), all of which are hallmarks of goal-directed behavior. Furthermore, the baseline model explicitly associates each net node to the corresponding brain subsystem. The policy system is implemented by the dorsolateral prefrontal cortex; the action system is implemented by the premotor cortex and the supplementary motor area. The state system is implemented by the medial temporal cortex, the medial frontal/parietal cortex and the caudate nucleus. Finally, the reward system is associated with the orbitofrontal cortex and the basolateral amigdala.

Similarly to most of RL goal-directed models, the baseline model assigns values to outcomes based on the current motivational state of the agent. When the motivational state changes, the utility function accordingly changes and new utility values are assigned. In this way, the agent is not able to anticipate the future motivational states. In next section we show how our model diverges from the baseline model

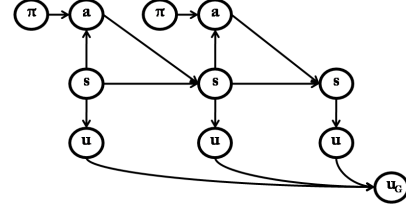


Figure 1: The Bayesian Model of Goal-directed Decision-making of Botvinick and An (2009).

in order to account for the ability of anticipating motivation. Afterwards, we compare the performance of the two models in three simulated experiments.

The model of anticipated motivation

The computational model that we propose (see fig. 2) accounts for the ability of anticipating motivational states by incorporating in the baseline model the explicit representation of the motivational system dynamics. During planning, the baseline model assigns utility considering only the current motivational state; rather, our extended model considers also future motivational states. In our model the agent knows that each future motivational state depends jointly on the previous motivational state and on whether (and to which amount) the agent has been satiated or not at the previous time step. In specific, in our model of anticipated motivation, state nodes are broken down in sub-nodes: *spatial states* (s), which represent the spatial position, *internal states* (i) which represent the motivational state, and *detection states* (d), which record the presence of potential rewards. Different motivation, such as hunger and thirst, have separated motivational state nodes and detection state nodes. For each motivation at a given time-step, the spatial state influences the detection state; for instance, if the food is in a certain (spatial) place, the agent has to be in that place to detect it. The detection state, jointly with the internal state, influences the internal state at the following time step. For example, at t_1 the agent is hungry (internal state) and it is near the tree (spatial state). Once the agent detects (detection state) and eats the food that is on the tree, at t_2 it is less hungry (following internal state).

As a consequence of the ability of anticipating motivation, during planning our agent assigns utility considering the relationships between potential rewards and motivations. Indeed, at each time step, utility u depends jointly on the motivational state i and on the potential reward detected d . Each motivation has its own associated utility node u . All utility nodes at all time steps are summed up by the global utility node (u_G).

In sum, compared to the baseline model, our agent plans considering also what we call the *motivational forward model*, which describes the dynamics of the agent’s homeostatic system (a system that monitors internal variables that are significant for the survival of the agent itself). It corresponds to the transition function linking both the internal value and the detection value at t_x with the internal value at

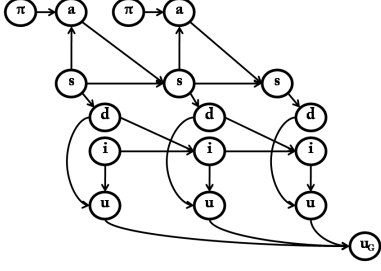


Figure 2: The Bayesian model of anticipated motivation.

t_{x+1} . In other words, the motivational forward model simply translates the following scenario in mathematical terms: if at t_x I am very hungry (internal state) and I see and eat a certain amount of food (detection state), then at t_{x+1} I am going to be less hungry (proportionally to the amount of food intake).

Experiments: method and results

We simulated our model of anticipated motivation in three simulated scenarios. In these simulations, we implemented a particular version of the general model of anticipated motivation described above, as shown in fig. 3. We considered an agent as characterized by two motivations: hunger and thirst. As said in the description of the general model, each motivational system has its own internal state nodes and detection state nodes. Thus, two internal state nodes are represented (*hunger* (h) and *thirst* (t)) and two corresponding detection state nodes, *food* (f) and *water* (w), respectively. At every time-step, internal node and detection node of each motivation jointly influence the corresponding utility, as described for the general model. Thus we have two utility nodes for each time step: u_H and u_T , for hunger and for thirst respectively. All utility nodes at all time steps are summed up by the global utility node (u_G).

Considering hunger as a paradigmatic example, “internal state nodes” can assume five values: 0, 1, 2, 3, 4 (0 means no hunger, 4 means maximum hunger). Similarly “detection nodes” can assume five values: 0, 1, 2, 3, 4 (0 means no food detected, 4 means maximum food detected). Spatial state values represent positions in a maze and can assume 5 values (corresponding to the maze positions). Action values are: “left”, “right” and “straight”. Policy values correspond to the combination between action and state values. The relationship between spatial states and actions (forward model) is different in each experiment and depends on the maze configuration (see each experiment section for details). The relationship between spatial positions and potential rewards changes in every experiment and it is described by the potential reward positions in the maze. The value of the internal state is the difference between the value of the internal state at the previous time-step minus the value of the detection state at the previous time-step; this accounts for the fact that hunger lowers by eating (as the same amount as the value of the eaten food). When the value of the internal state at the pre-

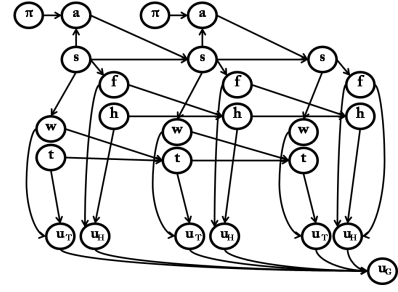


Figure 3: The model of anticipated motivation adopted in the simulations. It includes two drives, hunger and thirst, and two corresponding motivational forward models.

vious time-step is zero, the successive value is always raised by 2; this represents the increased hunger associated with the passage of time. Finally, the value of the internal state and the value of the detection state determine jointly the conditional probability of the utility corresponding to that motivation.

Like in the baseline model, utility is represented as the conditional probability of the binary variable $p(u/i, d)$. In simulations, we represented only positive or neutral utilities (being neutral utilities represented as $p(u = 1/i, d) = 0$ and maximum positive utilities as $p(u = 1/i, d) = 1$) associated to appetitive drives. In this model, at each time step $p(u = 1/i, d)$ corresponds to the minor value between detection state and internal state at that time state, over the maximum absolute value of the internal state (that corresponds to 4). For example, if potential reward detected is 2 and motivation is 0, then $p(u = 1/i, d) = 0/4$; if motivation is 1, then $p(u = 1/i, d) = 1/4$; if motivation is 3 and potential reward is 2, then $p(u = 1/i, d) = 2/4$.

The conditional probabilities of all nodes are deterministic, except $p(u/i, d)$. It means that, if the agent is in a position of the maze and makes a certain action, it will go deterministically to another given position. Similarly, if the agent is in a certain position and follows a given policy, it will always make a certain action; and if the agent is in a certain position, the potential reward in that position is always detected.

Experiment 1. Strategic planning

Humans and few other animals are able to inhibit prepotent responses, elicited by their actual motivational state, and to choose a course of actions that leads to higher rewards in the future. This is obtained by considering a complex prospect of future motivational states and corresponding future rewards and at the same time by exerting cognitive control over prepotent responses (Botvinick et al., 2001).

In our first experiment, we consider the context of a T-maze, as shown in tab. 1, left. We considered three time-steps: at t_0 the agent is in S_1 ; at t_1 it can go left to S_2 or right to S_3 ; at t_2 it goes from S_2 to S_4 and from S_3 to S_5 . In each of the five positions of the T-maze, a certain amount of food, water or both can be found. The configuration chosen

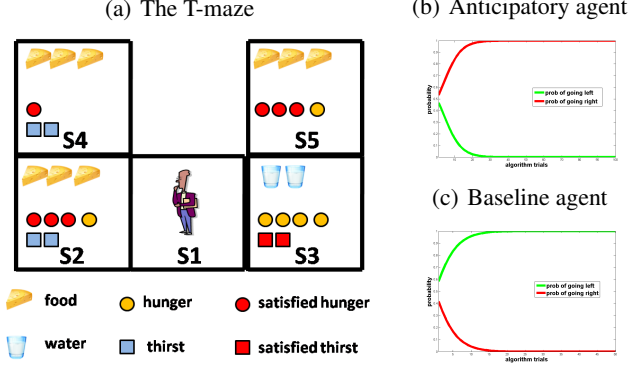


Table 1: Experiment 1. Left: T-maze. Symbols represent values of detection states and internal states that are computed by the anticipatory agent during the inference process. Potential reward pattern (corresponding to potential reward in each position of the maze) and initial motivational states (corresponding to motivational states in S2 and S3) are set by the experimenter, all further information is computed by reward query. Red forms indicate motivational values that are satiated by consumption of potential rewards in the corresponding position of the maze. Graphically, optimal behavior corresponds to choose the path with more red forms. Right: results of the first experiment (top = anticipatory agent; bottom = baseline agent). The graph represents the probability assigned to the policy associated to “going right” (red) and “going left” (green), respectively, at each iteration of the reward query.

in our simulation is the following: *food* 3 in S_2 , *water* 2 in S_3 , *food* 3 in S_4 and *food* 3 in S_5 . Then we set the initial internal states as follows: $H_1 = 4$, $T_1 = 2$. The critical feature of this set-up is the presence of two policies (going left or right) providing higher short-term and long-term cumulative reward, respectively. We hypothesized an advantage of the anticipatory agent in (long-term) utility maximization. In the anticipatory agent, utility is assigned to both current and future motivational states (related to both hunger and thirst). In the baseline agent, utility is assigned only to rewards that are congruent to the highest amongst actual motivational states of the agent (i.e. hunger). Tab 1, right, shows the results of comparison of anticipatory and baseline agents. In agreement with our hypothesis, in the case of the anticipatory agent the probability of choosing policy ‘going right’ increases monotonically towards 1 at every iteration of the reward query. The baseline agent has the opposite behavior, and chooses ‘going left’. Results indicate that the anticipatory agent is able to disregard the food that can be consumed immediately (by going left) in favor of a policy that maximizes its utility; indeed, by going right it satisfies both its thirst (at the second step) and its hunger (at the third step). On the contrary, the baseline agent, which considers only its current motivational state, acts impulsively and selects the policy that only gives short-term benefits.

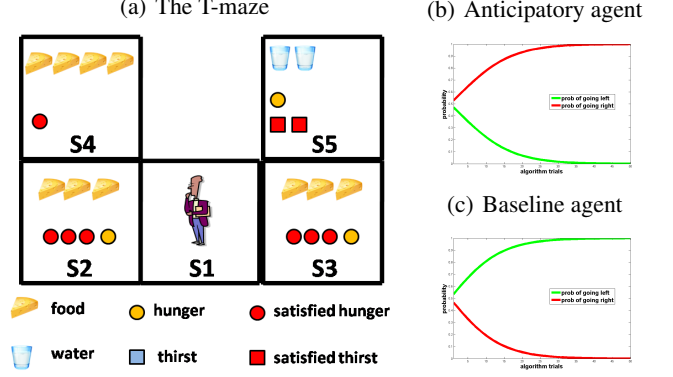


Table 2: Experiment 2. Left: T-maze. Right: results.

Experiment 2. Considering the future switch of motivation in the planning process

In a second experiment, we tested the agent’s ability to take its future changes of motivations into account during the planning process. The T-maze in tab. 2, left, illustrates this condition. Here potential rewards are: *food* 3 in S_2 and S_3 ; *food* 4 in S_4 ; *water* 2 in S_5 . We set the initial internal states of the agents as follows: $H_1 = 4$; $T_1 = 0$. A critical feature of this set-up is that it includes distal rewards (in S_4 and S_5) have different values if evaluated according to current or expected motivations. If a hungry agent ($H_1 = 4$) predicts that in the near future it will be satiated (i.e., it will collect food = 3), it can choose future potential rewards that at the moment seem lower (water = 2 rather than food = 4) but that will be higher (remind that, if at time t motivation is 0 it is raised by 2 at time $t + 1$). Similar to experiment 1, we hypothesized an advantage of the anticipatory agent. In agreement with our hypothesis, results (table 2, right) indicate that the anticipatory agent chooses to go right, and is able to satisfy both its (current) hunger and (anticipated) thirst. On the contrary, the baseline agent acts impulsively and prefers going left, because it cannot take into consideration that food of S_4 , when consumed, will no longer have high reward (because the agent’s motivational state is changed).

Experiment 3. Planning for the future: storing rewards in view of future needs

According to the Bischof-Kohler hypothesis (Suddendorf and Corballis, 1997), only humans, even if not motivated at the present moment, act in a complex and flexible way to procure rewards in view of future motivations (going, for instance, to the supermarket even if not actually hungry). Contrary to this idea, Raby et al. (2007) argued that even some other animals such as western scrub-jays (*Aphelocoma californica*) have this ability. In this work, experimenters thought scrub-jays to foresee conditions in which they would have received no food, thus feeling hungry; after this learning phase, experimenters gave to the scrub-jays the possibility to cache food. As a result, scrub-jays cached a larger amount of food when they foresaw a future condition of deprivation compared to

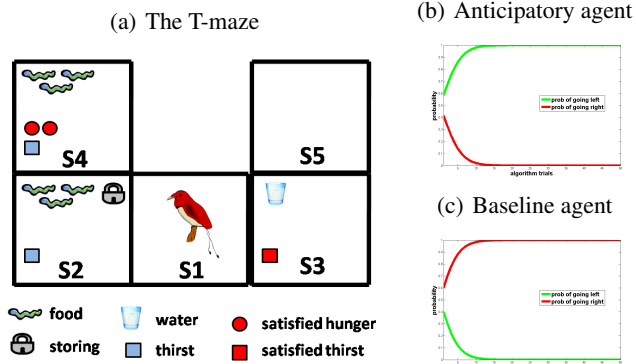


Table 3: Experiment 3. Left: T-maze. Right: results (top = agent with anticipated motivations; bottom = baseline model).

other conditions. These results may indicate that even western scrub-jays, at list to a certain extent, are capable to flexibly account for their future motivational states (although the results are controversial, see Roberts and Feeney, 2009).

Our third simulation is conceptually similar to the study of Raby et al. (2007) about the agent’s ability to store rewards in view of future motivational states (as shown in tab. 3, left). The scenario is again a T-maze: at t_0 the agent is in S_1 ; at t_1 it can go left (to S_2) or right (to S_3); at t_2 it goes from S_2 to S_4 and from S_3 to S_5 . Once a reward is found, the agent has two options: consuming it immediately or consuming it later, at the following time steps. Crucially, in our model of anticipated motivation, once the agent detects reward and, at the same time, it is not motivated, it automatically stores reward in view of future motivation. We positioned the following rewards: *food 3* in S_2 and *water 1* in S_3 , and set the initial internal state values to $H_1 = 0$ and $T_1 = 1$. Since in our model the agent can store reward if it’s not actually motivated, we hypothesized that the agent would have chosen to go left, storing *food 3* in S_2 (being not hungry) and consuming it in S_4 (once hungry; remind that in our model if a motivational state value is 0 at t_i it will be 2 at t_{i+1}).

Tab. 3, right, shows the results of our experiment. In agreement with our hypothesis, the anticipatory agent chooses to go left, storing a large amount of food and eating it later, instead of drinking immediately few water. In doing so, it inhibits the impulsive action of consuming the immediate reward (i.e., it does not choose water even if it is slightly thirsty) in order to obtain a higher reward in the future. On the contrary, the baseline agent acts impulsively; it is attracted only by the immediate reward, and is unable to plan instrumental actions that lead to the future consumption of larger rewards.

Conclusions

In this section, we have presented a Bayesian model of goal-directed behavior that accounts for future motivations during planning. Our model includes a component, a motivational forward model, for predicting future motivational states. Furthermore, utility values of outcomes depend jointly on reward

amount and on motivation at the corresponding time, while in most goal-directed RL they only depend on the former feature. While in most RL models motivations change utility function, in our model they are explicitly represented and thus influence the value of future potential rewards. Moreover, while most RL models assigns utility only in relation to a single motivation, usually the stronger, our model is able to integrate dynamics of many motivational systems.

We have shown that in three simulated scenarios, namely in the presence of a pressing impulsive need, in a future switch of motivation and when it is possible to store food, our model maximizes future rewards, contrary to RL models guided only by current motivation.

The debate on how human and non-human brain represents future motivations during planning is still controversial. Certainly, many animal species show behaviors that are prospectively oriented, namely functional to their future motivation. Migration, hibernation, nest building and food-caching are examples of such behaviors. Nonetheless, in most cases, these behaviors are possibly reactive. Indeed they can be in-built Pavlovian responses triggered by conditioned or unconditioned stimuli, or reinforced instrumental responses. For example, in Naqshbandi and Roberts (2006), squirrel monkeys could either eat four or just one date. Given that eating dates makes monkeys thirsty, experimenters manipulated the delay between the meal and the availability of water. In the one date case, water was available sooner respect to the four dates case. While at the beginning monkeys chose four dates, gradually they shifted their preference toward one date. However, as they did it after a long sequence of trials, their behavior was interpreted as reinforce-driven.

Conversely to this study, a goal-directed behavior has to be flexible. Being goal-directed behavior based on action-outcome contingency, it is immediately produced when it is likely that the desired outcome will be obtained. In a recent study, Raby et al. (2007) showed that scrub-jays cached food only when they expected future deprivation, and did it from the first trial. Similarly, Osvath and Osvath (2008) showed that chimpanzees and orangutans flexibly chose a tool for future use taking future needs into account.

Despite these studies suggest that, at least in some circumstances, some animals plan in view of future needs, Suddendorf and Corballis (2007) still consider human planning as unique. Indeed, they argue that the former is based on what they call *mental time travel*, which consists in mentally simulating, from a subjective perspective, past and future experiences in a vivid and flexible manner. This complex ability, related to episodic memory, would be the basis of planning, as it allows to generate rich future prospects. Some no-human animals could have similar prospection abilities, which could be implemented however using radically different neural mechanisms. In a similar vein, Raby and Clayton (2009) distinguish episodic and semantic systems of prospection; only the (more complex) episodic system might involve self-projection in the future.

In keeping with this view, at the neural level we hypothesize a qualitative difference on how human and non-human brains implement the *motivational forward model*. In keeping with the "mental time travel hypothesis", human vivid anticipation of future needs might partially activate brain structures associated to those needs. For instance, even if my homeostatic system does not currently need food intake, nonetheless thinking to the next Christmas lunch triggers my hunger. Specifically, the human ability of anticipating motivations during planning may depend on two interrelated brain processes. The first may be related to more abstract mechanisms of inhibiting preponderant responses and imaging future prospects, linked to areas such as dorsolateral prefrontal cortex and cingulate cortex. The second process may be linked to the activation of "as-if" motivations (Damasio, 1994) and hence may involve cortico-limbic structures directly related to motivations, like amygdala, orbitofrontal cortex, parahippocampal gyrus, and anterior fusiform gyrus (LaBar et al., 2001). The two processes may be connected as follows: cortical anterior structures may modulate the activation of cortico-limbic structures related to simulated motivations. The ability of simulating future needs in a vivid way, connected to the episodic memory and perhaps to the association between prefrontal and cortico-limbic structures, might be absent in animals. Indeed animals may be partially able to foresee thanks to rudimentary frontal processes. Furthermore, this ability might be impaired in patients with impulsivity problems. In this case, the deficit may depend on impairment in frontal lobe or in cortico-limbic structures. Designing experiments that test these hypothesis is an important avenue for future research.

Acknowledgments Research funded by the EU's FP7 under grant agreement no FP7-ICT-270108 (Goal-Leaders). We thank Matthew Botvinick for insightful comments.

References

- Balleine, B. W. and Dickinson, A. (1998). Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. *Neuropharmacology*, 37(4-5):407–419.
- Botvinick, M. M. and An, J. (2009). Goal-directed decision making in prefrontal cortex: a computational framework. In *Advances in Neural Information Processing Systems (NIPS)*.
- Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., and Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychol Rev*, 108(3):624–652.
- Botvinick, M. M., Ibara, S., and Prabhakar, J. (sub). Cognitive and neural foundations of goal-directed decision-making: An integrated framework.
- Buckner, R. L. and Carroll, D. C. (2007). Self-projection and the brain. *Trends Cogn Sci*, 11(2):49–57.
- Cooper, G. F. (1988). A method for using belief networks as influence diagrams. In *Proceedings of the Twelfth Conference on Uncertainty in Artificial Intelligence*, pages 55–63.
- Damasio, A. R. (1994). *Descartes' Error: Emotion, Reason and the Human Brain*. Grosset/Putnam, New York.
- Gilbert, D. T. and Wilson, T. D. (2007). Prospection: experiencing the future. *Science*, 317(5843):1351–1354.
- LaBar, K., Gitelman, D., Parrish, T., Kim, Y., Nobre, A., and Mesulam, M. (2001). Hunger selectively modulates corticolimbic activation to food stimuli in humans. *Behavioral Neuroscience*, 115(2):493–500.
- Murphy, K. P. (2002). *Dynamic bayesian networks: representation, inference and learning*. PhD thesis, UC Berkeley, Computer Science Division.
- Naqshbandi, M. and Roberts, W. (2006). Anticipation of Future Events in Squirrel Monkeys (*Saimiri sciureus*) and Rats (*Rattus norvegicus*): Tests of the Bischof-Köhler Hypothesis. *Journal of Comparative Psychology*, 120(4):345–357.
- Niv, Y., Joel, D., and Dayan, P. (2006). A normative perspective on motivation. *Trends in Cognitive Science*, 8:375–381.
- Osvath, M. and Osvath, H. (2008). Chimpanzee (*Pan troglodytes*) and orangutan (*Pongo abelii*) forethought: self-control and pre-experience in the face of future tool use. *Animal cognition*, 11(4):661–674.
- Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*. Cambridge University Press.
- Pezzulo, G. and Castelfranchi, C. (2009). Thinking as the control of imagination: a conceptual framework for goal-directed systems. *Psychological Research*, 73(4):559–577.
- Raby, C. R., Alexis, D. M., Dickinson, A., and Clayton, N. S. (2007). Planning for the future by western scrub-jays. *Nature*, 445(7130):919–921.
- Raby, C. R. and Clayton, N. S. (2009). Prospective cognition in animals. *Behav Processes*, 80(3):314–324.
- Roberts, W. A. and Feeney, M. C. (2009). Temporal sequencing is essential to future planning: response to osvath, raby and clayton. *Trends in Cognitive Sciences*, 14(2):52–53.
- Shachter, R. D. and Peot, M. A. (1992). Decision making using probabilistic inference methods. In *UAI '92: Proceedings of the eighth conference on Uncertainty in Artificial Intelligence*, pages 276–283, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.
- Suddendorf, T. and Corballis, M. C. (1997). Mental time travel and the evolution of the human mind. *Genet Soc Gen Psychol Monogr*, 123(2):133–167.
- Suddendorf, T. and Corballis, M. C. (2007). The evolution of foresight: What is mental time travel and is it unique to humans? *Behavioral and Brain Sciences*, 30(3):299–313.
- Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review*, 55:189–208.
- von Neumann, J. and Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton University Press.