A computational account of need states, reward and wanting

Bosulu, J[1]., Hétu, S.[1]

[1]Faculté des arts et des sciences, Université de Montréal

ABSTRACT

Needing is related to deprivation of something biologically significant, and wanting is linked to reward prediction and dopamine and usually has more power on behavioral activation than need states alone. The interaction and independence between needing and wanting in terms of brain functioning is not so defined, and it is unclear why sometimes there is association and sometimes dissociation between them. To investigate this question, we used a computational approach. In this approach needing is related to a deviation from preferred states that living things tend to occupy in order to reduce entropy, while wanting is related to precision over policy leading to rewards. We found that: (1) when need states increase, the tendency to occupy preferred states is enhanced independently of wanting (reward prediction), showing a dissociation between needing and wanting. We also found that (2) when need states increase, precision (i.e. saliency) of the reward leading to the preferred state is also increased, suggesting that need can amplify the value of a reward and its wanting. This led to the development of a model showing how needing is dissociated from wanting in that the tendency to be in a preferred state is independent of prediction of reward; and how needing can amplify wanting, by attributing saliency to relevant reward prediction.

INTRODUCTION

There has been a debate around the question of if people, for instance consumers, are driven by their needs or their wants (Campbell, 1998). Needing is related to a state of deprivation of something important for life or survival (Bouton, 2016), and increases arousal through interoceptive salience (Craig, 2003); and not responding to a need might lead to some adverse consequence such as physiological or psychological suffering which go beyond mere frustration (MacGregor 1960; Baumeister & Leary, 1995). Whereas 'Wanting' is related to the prediction of reward in the brain and is more closely related to dopaminergic activity and motivation (Berridge, 2004). Both have some influence, albeit differently, on the motivational value of stimuli. Indeed, motivation can have a directional effect and/or activational effect. The directional effect is linked to choice (preference or action selection) and directs towards or away from stimuli, while the activational effect is related to action and its initiation, maintenance, and vigor or effort (see Salamone et al., 2018). A recent fMRI meta-analysis compared needing

(perception of needed stimuli, i.e. food when hungry) and “wanting” (perception of a reward cue that leads to reward seeking). It showed that needing seems more related to brain regions implicated in directional motivation/value; whereas “wanting” seems to be more related to brain regions implicated in both directional and activational motivation/value, and to mesolimbic dopaminergic areas (Bosulu et al., 2022).

However, biological needs, such as those related to states of deprivation of a biologically significant stimuli or events (Bouton, 2016), seem to influence the rewarding aspect of biologically relevant stimuli, such as wanting, pleasure and choice. For instance food is often more wanted, liked and chosen when hungry. Thus, although it has been shown that motivation to pursue rewards depends more on expectations, i.e. reward prediction, than on needs states (Bindra, 1974; Berridge, 2004), need states tend to amplify reward predicting cues that are relevant to such need states (Toates, 1994; Berridge, 2004). Need states can also enhance liking of, and pleasure related to, relevant stimuli that satisfy one's needs (Cabanac, 2017; Berridge & Kringelbach, 2015; Becker et al., 2019). Moreover, a need state has the capacity to give and to control the preference/choice or value of a novel food or drink or of a particular choice or action, in relation to their consequence on the organism (Dickinson & Balleine 1994; Balleine, 1992), and such need related action happens via a system that is both dependent (Berridge, 2004; Berridge, 2018) and independent of dopamine (Wassum et al., 2011 Salamone et al., 2018). Even, in absence of a reward guiding cue, a need state can (directly) activate relevant actions (Passingham & Wise, 2012) or explorative behavior (Panksepp, 2004). This indicates that biological need states have an underlying effect through which they influence certain tendencies towards relevant needed stimuli that influences desire/wanting, liking, or goal directed behavior and choice.

The relationship between needing and wanting is of importance. It seems that in the brain they direct choice in a different way. Needing is related to interoceptive prediction, and especially interoceptive prediction error computed within the mid-posterior insula (Bosulu, et al., 2022; Livneh et al., 2020). Wanting is instead related to (exteroceptive) reward prediction and reward prediction error, with the latter computed within the ventral tegmental area (VTA) (Schultz et al., 1997; Bosulu, et al., 2022). Hence, needing seems related to the internal environment, whereas wanting seems to be more related to the external environment, although it also takes into account the internal environment (Berridge, 2004). Given the different implication of needing versus wanting in directional and activational motivation, and their relation with internal and external states, one can wonder computationally: what is the underlying effect that makes internal needs direct behavior (choice and action selection), even in absence of wanting? How does this same effect/directional action influence/amplify wanting (or liking)? In the following paragraphs, we try to answer these questions in a computational framework, using ideas and methods from active inference, information theory and reinforcement learning, and try to reconcile both questions computationally. More specifically, we will discuss how (and why) need states

directionally influence motivation and how (and why) this directional influence tends to amplify wanting (as well as pleasure, etc.). In other words, we will try to illustrate the mechanism of need states separately from wanting, in hope that this might explain the underlying mechanism of needing as well as how that mechanism allows needing to affect wanting, pleasure, interoceptive prediction, choice, etc., given the right context.

1 How needs influence directional motivation

The goal of organisms is to regulate internal states and keep them within certain very limited boundaries (Sterling & Laughlin, 2015; Friston, 2006). For instance, the average normal body temperature for humans is generally between 36.1°C (97°F) to 37.2°C (99°F), which is an incredibly small range compared to the range of possible temperatures in the universe, from the absolute zero to trillions of degrees. The same is true for levels of glucose or the balance between water and salt in the body. The main idea is that the number of states related to life are really small compared to the very large number of other combinations that wouldn't sustain life. So, in order for these limited boundaries to have a greater probability of occuring and for life to exist, natural mechanisms have set them as statistical priors, which are genetically encoded as innate preferred states. In that sense, living beings occupy very limited, preferred, states (physiological bound), which (from their point of view) have a greater probability, i.e. are less surprising (Friston, 2010). Here surprise (or surprising state), noted h, is the inverse of being probable and simply means less probable. Any self-organizing system must minimize such surprise (Friston, et al., 2006) in order to resist a natural tendency to disorder (Friston, 2010). Indeed, minimizing surprise allows living organisms to (temporarily) resist the second law of thermodynamics, which states that entropy - or the dispersion of a system - is always increasing. Entropy, noted H, is the average of the surprise (uncertainty), and it expresses the uncertainty related to which state must be occupied. It is in this sense that the presence of a state with high probability reduces entropy; and that such states can thus statistically be seen as priors (a priori) (Friston, 2010; Parr et al., 2022), or innate preferred states (that we will represent here by p) (Parr et al., 2022; Friston, 2010) as discussed above. It is also in this context that higher probability translates into more preference: living creatures naturally prefer those most probable, and limited, states that sustain life (Parr et al., 2022). Thus, distancing from those states leads to being in surprising states which are biologically costly and that generate entropy that living beings tend to minimize (Friston, 2010). Indeed, the notion of biological cost refers here to the fact that needing (i.e. being in surprising states) has a negative or biologically costly effect on the individual if s/he does not respond to that need (see MacGregor 1960; Baumeister & Leary, 1995). When a creature moves away from the preferred state, the tendency to occupy preferred states generates a state of need which translates into a “goal” of reducing entropy by going to (staying at) the states within physiological bounds, i.e the preferred state (Friston, 2010). Such tendency activates policies, i.e. state-action patterns, that compel creatures to seek out the (valuable) preferred

states. Here, we are using "policy" in a broader sense to mean both action or action pattern. With regard to need states, the policy is internal, i.e. it is generated by the internal environment as opposed to the external environment. The states that are in this policy path/trajectory take on a value (Friston & Ao, 2012). That is, by following a path (state-to-state policy) towards the preferred state, the intermediate states (situations, actions, stimuli) become valued, i.e. needed. For instance in going from hunger state to satiety (preferred) state, one intermediary state is having food (or eating); and such state would become valued because it is in the path to the preferred (satiety) state, as opposed to states that are not on this path. This is how a living creature would experience the need for food or the need to eat, and such stimuli or action will be preferred. In other words, the underlying effect of need states and the directional motivation related to them comes from the tendency to occupy preferred states.

The entropy discussed above is related to the surprising states that can only be changed by action, not perception, and represents the conjoint representation of all states in a given context, i.e. both surprising states and preferred states. It's in this sense that we argue that the presence/accessibility of, i.e. the possibility of going to, (limited) preferred state would reduce such entropy. We will refer to this "need related entropy" simply as entropy from now on, denoted H_N, and such entropy is given by: $H_N = H\ (h,\ \pi(p))$ (Eq 1); with H for entropy, $\pi(p)$ representing the path or policy π towards the preferred state (p), and h the biological cost (or value) of the (surprising) state. In this sense, the tendency to occupy preferred states bestows to need states their underlying influence on stimuli or actions when there is: 1) a conjoint representation of surprising states and preferred (less surprising) states (h and p); and b) a path towards the preferred state $\pi(p)$. In other words, in an environment where there are increasingly costly/surprising states, any stimulus/action that is in the path to the preferred state will become valued (needed) because it reduces entropy. In terms of perception, such value can be viewed as some form of internally generated saliency towards relevant stimuli.

2 How needs multiply wanting

The notion of precision helps understand the effect of needing on wanting (and on other phenomena such as pleasure). Mathematically, precision is a term used to express the inverse of entropy (Friston, 2010; Holmes 2022) and acts as a multiplicative weight on the prediction errors (change in prediction) (Parr et al., 2022). So, there are different precisions associated with different forms of changes in prediction (see Parr et al., 2022) (e.g. interoception, reward prediction, etc.) As discussed in previous paragraphs, it is the conjoint representation of surprising states and preferred states, as well as the accessibility of preferred states, that reduce entropy. In that sense, given increasingly surprising/costly states, any information that confirms a path to go to the preferred state, e.g. prediction of reward and/or the reward itself, will reduce

entropy, i.e. enhance precision, because the preferred state will have more probability of being occupied, i.e. there will be less uncertainty over which state to occupy. Indeed, such precision regarding reward signals (what to pursue?) or hedonic contact with the reward (is it good?) or the state of satiety (am I well?) might be respectively mediated by the dopamine system (Parr et al., 2022), the opioid system (Berridge & Kringelbach, 2015) and the serotonin system Parr et al., 2022; Liu et al., 2020; Luo et al., 2016). In this sense, when one is in a surprising state, the presence of a reward cue, i.e. a path to the preferred state, reduces entropy by improving precision via dopamine system activity (wanting); the presence of, or contact with, the reward itself reduces entropy by enhancing precision through the opioid system (pleasure/liking); while being directly in a preferred state or towards the preferred state could be related to precision via serotonin (well-being). All of these contexts follow the logic that they serve as information that reduce entropy by signaling (i.e. being in) a path to preferred states ($\pi(p)$) in opposition to surprising states (h), as given by the relation: $H_N = H(h, \pi(p))$.

Focussing on wanting, the neuromodulator dopamine, which is linked to prediction of reward and behavioral activation (Hamid et al., 2016), is said to represent the precision over rewarding policies (incentive salience), i.e. the confidence that a policy will lead to reward (Parr et al. 2022, Holmes, 2022). Here the policies are external, i.e. depends on the environment that is external to the animal. Also, the policies can be emitted either by the individual or by the environnement, and can be viewed as a sequence of events. Thus a highly precise belief signals, i.e. confidence, that a policy will lead to reward will lead to increase in dopamine signaling. If that reward that can be expected to minimize need states (biological cost), the policy leading to such reward would be reflected in even more dopamine signaling (FitzGerald et al., 2015); so that reward cues that increase the precision of beliefs about policies trigger dopamine bursts, which will attribute incentive salience to such stimuli (Berridge, 2007). Indeed, wanting depends on external stimuli that act as pavlovian cues that predict rewards (Berridge, 2018), and the attribution of value to these cues depends on mesolimbic dopamine which can be enhanced by physiological states (needs, emotions, drugs, etc.). (Berridge, 1996; Berridge, 2004). In that sense, Zhang et. al. (2009) provided a computational model for incentive salience 'wanting', suggesting a physiological variable, k (more specifically representing dopaminergic state), that changes the value of reward, r, that contributes to the value of states, V(s), resulting in an amplified value where $\overline{V}(s_t) = \overline{r}(r_t, k) + \gamma V(s_{t+1})$. (With γ representing a discount factor, t for time step, and $V(s_{t+1})$ the value of the next state). Hence, the value of a reward cue depends on the synergy between the reward value r and the dopaminergic states k. If r can be a proxy of reward that assigns state values, and *k,* is influenced by the physiological need states, and in this case surprising state (hunger or thirst, etc) then $\overline{r}$, which is the function of the combination of r and k reduces entropy. In this sense, the state value $\overline{V}(s_t)$ is enhanced because of both the physiological need state k (surprising state) and the value of the reward r, that signals the path to

the preferred state (p). The amplified variation of reward is a function of the reward itself and k, that is: $\bar{r}(r_t, k)$. Here the reward is (on) the path to the preferred state. Hence, replacing the path/policy to preferred state, that is the reward r, by $\pi(p)$, and the physiological state k by the surprising state h, we have that the function of the enhanced reward $\bar{r}$, which is encoded as confidence by dopamine, is related to precision ; that is the inverse of need related entropy, i.e. inverse of the function $H_N = H\ (\pi(p),\ h)$, or in other terms: the function $\bar{r}(r_t, k) = H^{-1}\ (\pi(p),\ h) = H_N^{-1}$; showing a direct equality between our model and Zhang's (2009) model. Our model is different and allows a more general treatment in that it expresses that relation in terms of information and uncertainty. This means that by signaling precision of policies leading to the needed reward, the precision associated with reward cues that generate wanting is enhanced as they reduce entropy, .i.e they reduce the uncertainty related to which state to go to ; and they do so because they lead to the reward that leads to the preferred state. In this sense, enhanced precision means more certainty that the state to which the policy leads will certainly be rewarding, and it is this certainty that amplifies wanting. For instance a relatively satiated animal that senses by smell the probability of food, will be motivated to follow the smell because there is probability of food. However, an hungry animal that senses by smell the probability of food will be motivated to follow the smell, not only because there is probability of food, but even more because that food will be very rewarding. That is the animal will be more certain (i.e. there will be more precision) that it should pursue the reward being predicted. It's in that sense that needing amplifies wanting. We can also imagine that, for the same reasons, the hungry animal will likely have more pleasure while eating food than the satiated one, i.e. more certainty (precision) will be assigned to how good food is when one is hungry. Hence this need related precision enhancement can be viewed as need generated saliency, i.e. need states make salient: rewards, stimuli or actions that lead to the preferred state.

That being said, the directional and underlying influence of need states separately from reward prediction, i.e. wanting, has not been modeled. In the same sense, there is no model for how this same underlying influence of need amplifies wanting, i.e. increases the precision of reward cues that lead to the preferred state.

HYPOTHESIS

Our first hypothesis addresses the directional influence of need states. Specifically, need states gain their underlying effect from the tendency to occupy preferred states. When one moves away from such preferred states and onto costly states, a state of need is generated. In that sense, costly or need states increase the probability of activation of policies that lead to preferred states ; and the conjoint presence of an increase in biological costs (i.e. increase in needs) and of a path to preferred states reduces entropy. In that sense, we hypothesize first that the deprivation/biological cost of 'needing' will have more influence on the probability, i.e. the

tendency towards the rewarding state, than simply reward prediction (as computed by the agent in frameworks such as reinforcement learning, where they try to maximize rewards), which can be related to wanting (Zhang et al., 2009; Berridge, 2004). This will show the directional and underlying effect of needing, separately from wanting, i.e. separately from reward prediction. The second hypothesis addresses the fact that, given the right context, need or costly states influence values of relevant stimuli because those stimuli reduce entropy. Specifically, the presence of both increase in cost of the surprising state as well as presence of a path/policy to the preferred state, will reduce entropy, i.e. enhance precision. That is, increasingly surprising states will lead to increased precision (reduced entropy) if there is a path to the rewarding state, because one will be more and more confident about what to do (or how good this is), i.e. what state to go to (or stay at). As wanting is related to cues that signal precision over policies leading to reward, this will show how need states amplify wanting by enhancing precision of reward cues, if the predicted rewards lead to the preferred state.

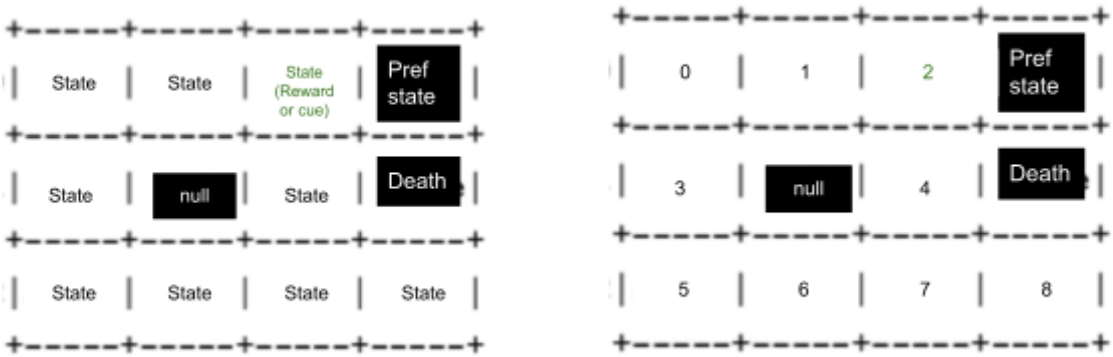
METHOD

Computational model

Because many ideas developed here stem from the Active Inference framework (see Parr et al., 2022), informational entropy, and the free energy principle (see Friston, 2010), and to avoid tautological testing (testing the theory using the tools provided by the same theory), we will use a different framework to illustrate the most important ideas and hypotheses developed here. Our hypothesis stem from mathematical properties and would likely apply in different computational frameworks. Indeed, one way to test our ideas computationally is through the reinforcement learning (RL) framework, specifically because it involves an agent that has a goal and is based on policies over state transitions to achieve such a goal (Sutton and Barto, 2018).

In order to mimic the idea of limited physiological bound, we created an environment where only one state will be a rewarding/preferred state and where the number of surprising/costly states is disproportionately higher, and . We can draw a parallel with the example of human hunger or temperature: the preferred state corresponds to the optimal interval of sugar level in the bloodstream, or the temperature range (between 36.1°C (97°F) and 37.2°C (99°F)). In this example, each interval of sugar level, or of temperature range, would be related to a specific state. This can be represented by a grid-world, with different boxes, representing the states in which one can be. Hence a 3x4 grid-world type environment containing 9 states (fully white boxes) to be valued was used. The other states were two terminal states: one being the rewarding/preferred state, with a reward of 1, and the other terminal state simply had a negative reward of -1 (death). One state (state between 3 and 4) was inaccessible and neutral (a kind of wall) with a reward and a value that remained zero (see fig. 1).

Fig. 1

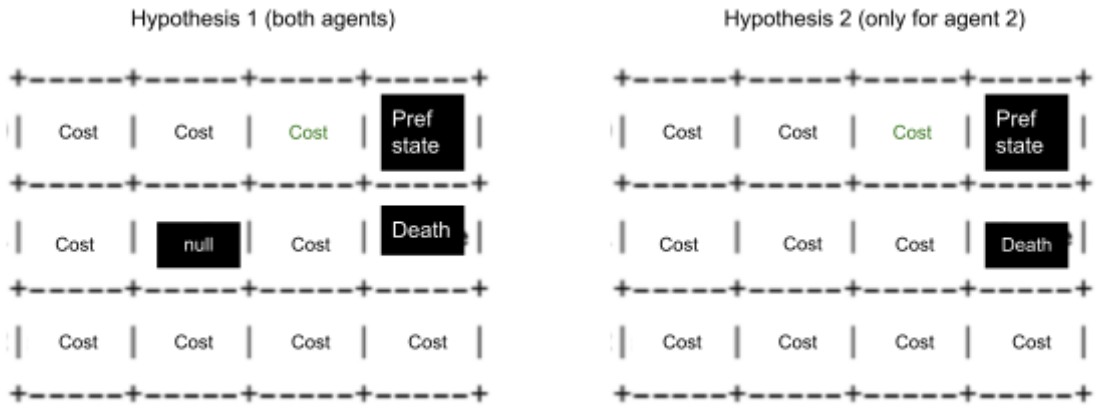


Grid world representing different states as well as the preferred state and the reward or reward cue that leads to the preferred state. (The second grid shows the states numbered which is useful in order to understand the results for the first hypothesis.) Here the agent will have to attribute either value or probability to states ; and the state 2 is the one that is the most on the path the the prefere state.

In order to illustrate our hypotheses, we will vary only the amount of cost in costly states. That is, we are testing how need/costly states influence tendency towards the rewarding/preferred state (hypothesis 1), and how and if the presence of a path to the preferred state in a more and more costly environnement results in high precision and reduced entropy (hypothesis 2). In order to add biological cost (i.e. “deprive” the agent), we “charged” all states with a biological/deprivation cost, i.e. a negative reward representing a significant departure from biological bounds, which were either -1, -10, or -100 for all states, except the terminal states and the wall for the first hypothesis (see fig. 2). For the second hypothesis all states, except preferred state, were embedded with the same biological costs (the wall was removed) (see fig. 2). Of note,

the (biological) cost is defined by us, but (all) the agents compute the *expected* (biological cost), i.e. Q values, following the formula described further below.

Fig. 2



Grid world representing different states for hypothesis 1 and 2. In hypothesis 2 all states bear the same cost except the preferred state. In both hypotheses, the state 2 (in green) is still the one that is the most on the path to the preferred states.

Across all experiments, the hyperparameter γ gamma, which controls the discount of options (states), was set to 0.9, the learning rate α was set to 0.1, and we used epsilon greedy of 0.1, for the exploration vs exploitation balance, which are standard in RL (see Sutton & Barto, 2018). Agents learned their values over 1,000 iterations.

To test the first hypothesis, that need/costly states influence much more tendency towards the rewarding/preferred state, as compared to reward prediction, we compared two agents: one that computed action values and another that further implemented tendencies/policies. In order to fairly compare the two agents, they both used the same computation for the learning part agent and valuing actions, but agent 2 did not use those values for decision, but went further and turned them into probabilities/tendencies. (we should note that the results would be the same for agent 2 even without using the same learning part as agent 1).

Hypothesis 1

Agent 1 and agent 2

The first agent make decisions based on state-action values of each decision given the current states (see Sutton & Barto, 2018). Here the policies depend on the action values, denoted $Q^{\pi}(s, a)$, and given by:

$$Q^{\pi}(s,a) = E_{\pi}\{R_t | s_t = s,\ a_t = a\} = E_{\pi}\{\sum_{t=0} y^i r_{t+i} | s_t = s,\ a_t = a\} \quad \text{Eq. 1}$$

The equation shows the value or "quality" (Q) of the action (a) in state (s) under a policy (π). The function denoted $Q^{\pi}(s,a)$ expresses the expected (E) return (R) starting from state (s) and taking the action (a), and thereafter following policy (π). The agent's decision after learning is based on the optimal policy π_*, i.e. the one that maximizes the expected return, and therefore the optimal $Q^{\pi}(s,a)$, noted $Q^*(s,a)$ is equal to:

$$max_a Q(s,a) \quad \text{Eq. 2}$$

where max_a is related to the action that maximizes $Q(s,a)$.

It has been argued that relative values are an important measure of the value of a choice (see Schultz, 2015). Therefore we builda second version of agent 1 which will also be able to compute values in a relative way, i.e. make decisions based on relative values derived from those in the equation above, as follows:

$$Q_{rl}(s,a) = Q^*(s,a) - \mu(Q^*(s,a)) \quad \text{Eq. 3}$$

Where $Q_{rl}(s,a)$ represents the derived relative values of taking action a in state *s*, $Q^*(s,a)$ represent the state-action pair with the maximal value in state *s*, and μ is is the function for the mean; in other words, the $Q_{rl}(s,a)$ is the state-action value relative to the mean of maximal action values. However, our hypothesis 1 is different from relative value in the sense that it normalizes the values (using a softmax function as shown further below).

Agent 2

Agent 2 went further and used the Q values to compute the tendencies. In order to test our first hypothesis, that concerns how need states direct action/motivation, the second agent will use a probability (i.e. preference or tendency) P(s,a), for action a, in state s, where $P(s,a)$ is somewhat equivalent to a probabilistic version of $Q^*(s,a)$. (Here uppercase *P* indicates probability and is different from lowercase *p*, which is used here as the preferred state.) In other words the agent 2 uses the same action value function $Q^*(s,a)$ and takes things further by turning that action value into $P(s,a)$.

Indeed, mathematically preferences can be assessed using a softmax function which bounds the values between 0 and 1 and gives them the probability according to their scale. Assuming we have a context (vector), containing all states (i.e. state-action pairs), both the surprising and preferred ones, with different categories of policies, the probability of a category of policy will be activated is given by:

$P(s,a)) \ = Cat\ (\pi_0)$ Eq. 2 (see Parr et al., 2022)

This means the probability (P) of a tendency to do action a from state s, is determined by the category (Cat) parameterized by π_0

The parameter π_0 itself is a softmax function over optimal action values ($Q^*(s,a)$) in the form:

$\pi_0 = \sigma\ (Q^*(s,a))$ Eq. 3 (see Parr et al., 2022)

where σ represents the softmax function, bounded between 0 and 1, and enforces normalization (i.e., ensures that the probability over policies sums to one (in the given context).

Our first hypothesis is that $P(s,a)$ will be higher for the preferred state (which is rewarding and whose value is 1); and that *P(s,a)* will be lower for (biologically) costly/surprising states, i.e. those with a negative value. Because only agent 2 calculates such preferences, we expect to see more impact of biological costs on agent 2's results than on agent 1's, as s/he only computes and uses the Q-values. This would illustrate the directional value/motivation related to need, i.e. the tendency/probability to go to (or stay at) the preferred state, separately from reward prediction (as discussed the same results would happen for agent 2 even without computing the Q-values). So, the assumption is that, the more the biological costs increase (state of need), the preference for the reward state (i.e. the state that leads to the preferred state) will be much higher or more important for agent 2 which uses tendencies compared to agent 1 which just uses reward prediction, i.e. state-action values (absolute or relative).

Hypothesis 2

Our second hypothesis is about how the conjoint presence of costly (surprising) states and that of the preferred state, as well as a path towards it, reduces entropy and thus increases precision (which can be viewed as a need generated saliency). This results in neural activity that enhances wanting, liking, etc. depending on the context. In order to test our second hypothesis, we will

look at the change in precision (inverse entropy) that comes from increased biological cost (surprise) when there is a (path to) reward versus no reward. Specifically here we test wanting, but the logic would apply to other processes (e.g. liking).

The entropy of a system is given by:

$$H = \sum - \ln(P(s,a) \times P(s,a))$$

Where H represents entropy, i.e. the average surprise or the uncertainty over which states to occupy, and $-\ln(P(s,a))$ represents surprise $h(s,a)$ of being in a state s and taking action a. Hence the equation can be written:

$$H = \sum h(s,a) \times P(s,a))$$

In other words H is a function of $h(s,a)$ and $P(s,a)$, and following the usual mathematical expression of a function: f(x), the above can be written $H(h, P(s,a))$.
Because $P(s,a)$ depends on the parameter π; this can be written : $H(h, \pi_0)$, to emphasize that the entropy is a function of surprising states and the parameters that determine their probability.

As discussed, need related entropy, H_N , is reduced when there is a path towards the preferred state $\pi(p)$, when there are costly/surprising states. Thus the formula can be written as in the introduction:

$$H_N = H(h, \pi(p))$$

Defining H_N as the uncertainty over which states to occupy means that surprising states increase entropy, but the presence of a preferred (non surprising) state will decrease entropy if there are surprising states. In other words, if the surprising states are more and more costly, and there is a path to the preferred state, the creature/agent will be more certain about which state to go to, hence the reduction of entropy and enhancement of precision. Mathematically, the probabilities of increasingly costly/surprising states will go further towards 0, while going further towards 1 for the preferred state, which will have much higher probability and thus entropy will be reduced. In other words, our hypothesis is that the conjoint presence of costly (surprising) states and that of preferred states, i.e a path towards them, reduces entropy and that is why need (biologically costly/surprising) states increases precision of reward and reward cues, as they lead to preferred state. The precision is given by:

$$P_N = 1/H_N$$

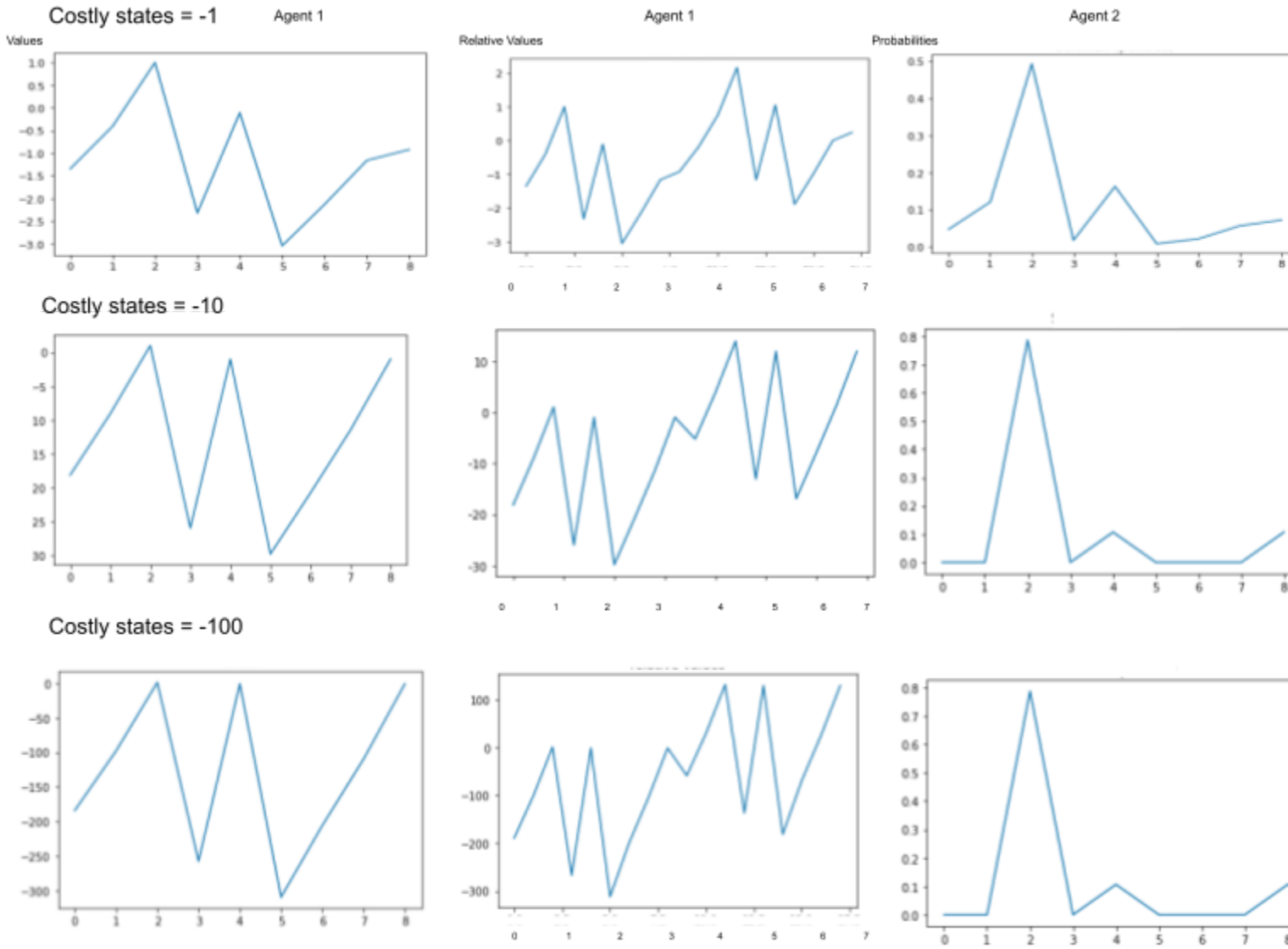
Where P_N is for need related precision.

RESULTS

We hypothesized first that the deprivation/biological cost of 'needing' will have more influence on the probability, i.e. the tendency towards the rewarding state, than their absolute or relative values. Second, the conjoint presence of the more costly/surprising state and the path/policy to the preferred state, will reduce entropy, i.e. enhance precision. That is an increasingly surprising state will lead to more probability of preferring the preferred state; and will thus lead to increased precision (reduced entropy) if there is a path to the rewarding/preferred state, because one will be more and more certain about which state to prefer, and thus confident about what to do, i.e. what state to go to. The first hypothesis will represent the directional influence of need states and the second will represent the fact that need states influence value relevant reward cues because those cues reduce entropy.

Our first hypothesis is that $P(s, a))$ will be higher for the state that leads to the preferred state (state 2); and lower probability for (biologically) costly/surprising states. This would illustrate need related directional value, i.e. the tendency to go to the preferred states, when there are, or we are in, need (i.e. costly) states.

We show that expected biological cost influences the tendencies of behavioral patterns more than action values. That is, biological costs do not control or constrain values directly, but directly



control probabilities, i.e. policies or tendencies, to be in (or go to) the preferred states.

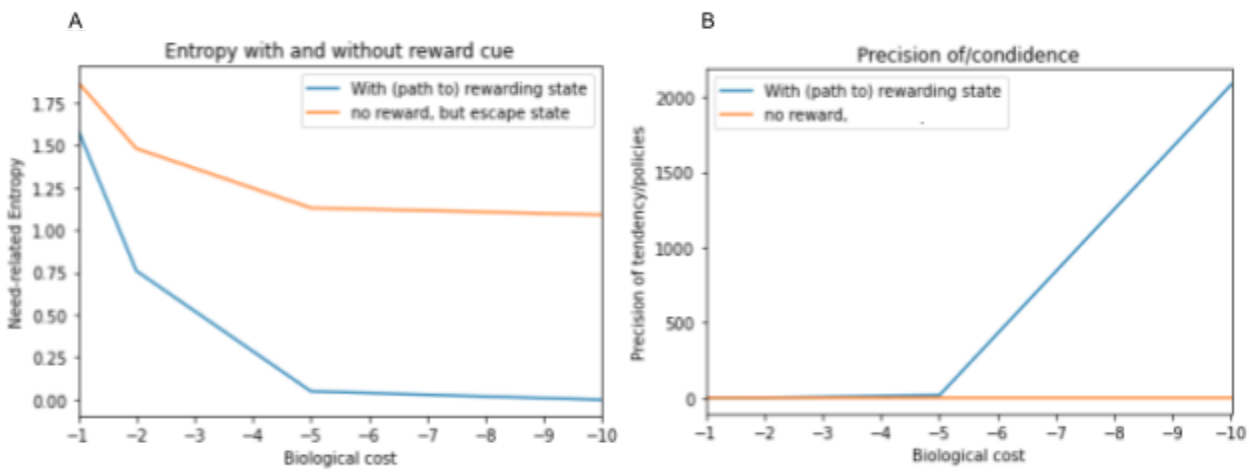
Fig 3

The results show that increasing (biological) costs have more impact on the probabilities of the only rewarding state (state 2) leading to the preferred state (computed by agent 2), as opposed to other states, whereas this effect is absent when looking at reward prediction related values (computed by agent 1). This is consistent with our hypothesis that need (directionally) influence

tendencies (i.e. probabilities) rather than values. Such results are due to the fact that agent 2 transforms the initial values of states into preferences in terms of probability, and as the expensive states are numerous (and costly/negative), the state containing the reward becomes very probable; which is not visible in agent 1 which only calculates (action) value.

Our second hypothesis was that the need-related entropy, $H_N = H\ (h, \pi(p))$, will be reduced/low with increasing biological costs if there is a path to the preferred state as compared to no real preferred state. Such entropy reduction will be related to increased precision. The logic being that, because entropy represents the uncertainty about what state (is more probable) to occupy; the more costly the surprising states, if there is a path to the preferred state, the more the probability of the creature/agent going to (or staying at) the preferred state ; hence the reduction of entropy and enhancement of precision.

Our results show that conjoint presence of increasingly (biologically) costly states and path to preferred state reduce entropy and increase precision (certainty over which state to occupy), and that the more the surprising states are costly, the more the agent is certain that he/she wants to be in, or to go to, the preferred state. (Here no reward state means all states are (biologically) costly, and hence there is no real preferred state.)



Graphs showing the impact of biological costs (same cost for all affected states as described in method) on the entropy (graph A) and on the precision/confidence (graph B), when there is a reward, i.e. path to the preferred state compared to no reward/no preferred state. Entropy reduction (or increase in precision) means that there is more probability that the agent will occupy the less costly, i.e. the preferred, state; given the increasing costs on other states. In this case, the more the biological costs, the more the agent is certain about in what state to be or to

go to. (Precision increases faster when entropy gets closer to zero ; and because it goes to 2000, it is hard to see the increases between -1 and -5 because of the scale it later reaches (2000). Also the -100 is not shown as the precision goes incredibly higher). It should be noted however that the reduction of entropy in the no reward is due to the fact that it is still better for the agent to go to the terminal state, i.e. to go to "death", than stay in the environnement with increasing biological costs.

DISCUSSION

Our goal was to look at the computational way in which need states enhance tendencies that direct actions and choice, and result in enhancement of wanting (and liking). Our first hypothesis concerned how need states direct action/motivation: for living creatures, there are very limited physiological bounds, i.e. preferred states, as opposed to surprising states, and living creatures have a tendency to go to or stay at the preferred states. Need states are related to biological costs of being in surprising states, and enhance the tendencies or policies that lead to the preferred state. The second hypothesis was that, if one takes into account both the limited preferred state and the surprising states, the entropy would be reduced and the precision enhanced the more the biological cost on the surprising states, if there is a path to the preferred state. To test this, we used the reinforcement learning framework, which involves an agent that has a goal, and finds policies over state transitions to achieve it. In order to test the first hypothesis, we compared two agents that are making decisions, one by attributing either absolute or relative action values to states, and the other by attributing probabilities of state action policies, i.e. tendencies. For the second hypothesis, we measured the variation of entropy or precision with regard to increasing biological costs of surprising states when there is a path to reward versus when there is no reward. With regard to the first hypothesis, we show that biological costs influence tendencies (probability of policies), positively towards the preferred states; as shown for the agent 2 that computes such probabilities using a softmax function, compared to the agent 1 that only computes (relative or absolute) action values. This explains the directional effect of need states, in the sense that states within the path/policy trajectory that lead to preferred states become rewarding and preferred as so it can direct motivation. Concerning our second hypothesis, our results show that the conjoint presence of biologically costly states and preferred state leads to a more reduced entropy and an even greater increase in precision, and this can be viewed as the more the costly (surprising) states are, the more the agent is certain that he/she needs to go to the preferred state. This explains how need states amplify wanting or liking of stimuli: they reduce entropy or more clearly, they precise the path/action to take/do.

Need related directional motivation

Our results confirmed our hypothesis and showed that the presence of the biological costs (surprising states), i.e. need states, influence the tendencies (probability of policies) positively towards the preferred states and negatively for the surprising/costly states; and this tendency (computed by agent 2) is not dependent on the prediction of reward (computed by agent 1). This result could be used to explain the directional and underlying effect of need states: need controls directional motivation, because of the tendency of living beings to move towards preferred states ; and such tendency activates policies that lead to the preferred states. In doing so, states within the trajectory of those policies become preferred, i.e. valued, in the presence of costly/need states. This tendency is mediated by homeostasis or its more general form allostasis (Sterling, 2004; Barrett, 2017; Holmes, 2022; Demekas et al., 2020), processes that help animals reverse or resist entropy (Sterling, 2004; Holmes, 2022) both physiologically and behaviourally (Holmes, 2022). In other words, preferred states, likely evolutionarily defined (Friston, 2010), are maintained by homeostatic or allostatic regulation via behavior (Parr et al., 2022), at the somatic, autonomic and neuroendocrine levels (see Swanson, 2000) ; and act as drives or goals that, through homeostasis and allostasis, direct action (Barrett, 2017); hence the directional effect of need states. Such directional influence, dependent on tendency to occupy preferred states, is also responsible for the underlying effect through which need states amplify wanting, pleasure, interoceptive prediction, choice, etc.; by enhancing precision of their related stimuli (or actions) that are in the path towards the preferred state, in an environment of costly/surprising states. This will be discussed below.

The path towards the preferred state

The second simulation showed that when the surprising states (need state) that create entropy become more and more costly, this entropy is more reduced, and precision is much more increased, if there is a reward or a reward signal that indicates a path to the preferred state, compared to if there is none. This confirms our hypothesis that the formula of need related entropy reduction is conditional to a path towards the preferred state. In that sense, any stimuli, cue, or action that is on such a path will reduce need related entropy, and enhance precision because the animal or his body will know what to do to go to the preferred state. It should be noted however that in our results, there was a slight reduction of entropy when there is no reward. This is due to the fact that it is still better for the agent to go to the terminal state, i.e. to go to “death”, than stay in the environnement with increasing biological costs. However, when there was a reward, i.e. a path to the preferred state, the reduction of entropy and especially the increase in precision were still higher. This might further explain the underlying effect of need states on actions and stimuli that lead to the preferred state. When in need states, the precision of relevant stimuli and actions is increased as they reduce entropy ; and such increase in precision can be viewed as a need generated saliency. In that sense, our result might explain how need states amplify wanting, i.e. amplify value of reward cues. We know that wanting is related to

both reward cues that activate dopamine, which is linked to precision of a policy leading to reward (Parr et al. 2022, Holmes, 2022), and to physiological states such as need states. In that sense, need (surprising) states can amplify such precision signaled by reward cues, i.e. make them more salient (e.g. via need induced higher dopamine reactivity), because they lead to preferred state and reduce entropy.

Needing vs wanting

Needing and wanting can happen both together and separately (see Bosulu et al., 2022). First need can happen without wanting, because wanting relies on reward prediction and our first results show that need is related to the tendencies to occupy the preferred states, and can influence preference of reward that lead to preferred state even when such influence is relatively low or absent in reward prediction frameworks. When needing and wanting for a stimulus (or action) happen together, need states tend to amplify wanting (Berridge, 2004 ; Toates, 1994) by increasing precision, or saliency, of wanted cues as shown in our second results.

Furthermore, as discussed there are different precisions associated with different forms of changes in prediction (see Parr et al., 2022). So our result lets us speculate that needing can increase different forms of precision, or salience, such as those related to wanting, i.e. “incentive salience” (see Berridge 2004), but also to liking or pleasure, etc. as long as they are on the path to the preferred state. In the same way, needing increases precision of stimuli or actions only if they are available and fulfill their role. For instance, for need states to amplify pleasure, the agent must come into contact with the reward. In the same sense, for the need states to amplify wanting, there must be the presence of a cue that predicts that the policy (sequence of events) will lead to reward. In that sense, Bosulu and colleagues (2022) conducted a fMRI meta-analysis on perception of needed stimuli in absence of wanting, and found that need states did not sufficiently and consistently activate the dopaminergic system, which is related to wanting. This might be due to the fact that the perceived needed stimuli did not sufficiently indicate that, in the sequence of events, they would be obtained ; and need states can not enhance precision of nothing. It's in this sense that needing can happen independently of wanting. However, the same study showed consistent activity within the mid-posterior insula (Bosulu et al., 2022), which might be interpreted as that needing can amplify the precision regarding change in interoceptive prediction, even in absence of a “true” reward prediction (see Bosulu, et al., 2022). Also, wanting can happen without needing as it is generated by reward prediction, which is related to precision over policy and is technically independent of need (Bindra, 1974; Berridge, 2004). In addition to this, the wanting can be amplified by internal states other than the need states, such as emotions, stress, drugs, etc. (Berridge, 1996; Berridge, 2004). This likely means that those non-need related internal states can amplify wanting related precision, and that need states are perhaps not the only ones to generate entropy or influence the reactivity of the dopaminergic system. Thus there can be needing without wanting, and wanting without needing, and there can both happen synchronously.

Furthermore, our study showed that tendencies could be much more influenced by need states than reward prediction suggests, and that in case of need states the policy comes first and determines reward. Whereas wanting, being related to reward prediction, is determined by reward. This difference in precedence between policy and reward may be due to the fact that in the case of need, it is an internal policy, related to the functioning of the organism, possibly through homeostasis or allostasis. While wanting is related to external policies, i.e. changes in the environment related to prediction of reward which can be attained by the behavior (likely related to the somatic motor system (see Swanson, 2000). This is one of the reasons why wanting has more control over the activation of behavior portion of motivation than needing (Bosulu et al., 2022; Salamone et al., 2018; Berridge, 2004). Moreover, this also gives another perspective on our second results which showed that need states increased precision when there is a reward leading to the prefered state. In that sense, if the reward determined by the need states, i.e. by the internal policy, is the same as the reward predicted by a signal of reward which specifies the external policy; then the precision will be enhanced and needing and wanting will be synchronized.

Limits and strengths

This study went further in abstraction to try to define the difference and interaction between needing and wanting, and the underlying effect of need states on wanting and other phenomena such as pleasure, preference, etc.. However, the notions discussed here might be difficult to find or define in the brain; for instance what we define as the “need related entropy”. Also, even the notion of biological cost, although interesting, has limits: are all needs linked to a biological cost if they are not met? We can think of needs related to reproduction, for example. Thus, this paper has several limitations that could be addressed in future research.

Nevertheless, the hypotheses discussed in this paper are mathematically established amd rigorous, and they are independent of a particular framework; and this was shown as we used a reinforcement learning framework, although some of our key ideas come from other frameworks, such as Active Inference (see Parr et al., 2022), informational entropy, etc.. Moreover, the hypotheses discussed have grounded roots in brain and behavior related literature. Also, the paper introduces the notion of, and explains, the underlying effect of the need states,. This idea leads to the suggestion of the distinction between internal and external policy. This latter idea makes it possible to explain why wanting has more control on behavioral activation (Berridge, 2004; Salamone, 2018), but that need can modulate wanting (Berridge, 1996 ; 2004) just as it can modulate liking (Berridge, 2007), preference (Balleine, 1992), etc. trough need generate saliency that we described here as need related precision. This also explains the directional effect of need and both the activational and directional effect of wanting (see Bosulu et al., 2022). Furthermore the notion of “need related entropy”, and the possibility that there are other phenomena which

could cause entropy and excite amplify wanting explains how dopamine reactivity such as related to stress, emotions, etc. can amplify wanting (Berridge, 1996; Berridge, 2004).

.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

ACKNOWLEDGMENTS

The research was supported in part by NSERC Discovery Grant #RGPIN-2018-05698 and UdeM institutional funds.

AUTHOR CONTRIBUTION

Juvénal Bosulu: Designed the study, performed the experiment, data analysis, interpretation, and wrote the manuscript. **Sébastien Hétu**: revised the manuscript and provided critical feedbacks. All authors contributed to and approved the final manuscript version.

DATA AVAILABILITY STATEMENT

All Data are available upon request.

REFERENCES

Balleine, B. (1992). Instrumental performance following a shift in primary motivation depends on incentive learning. *Journal of Experimental Psychology: Animal Behavior Processes, 18(3),* 236.
https://doi.org/10.1037/0097-7403.18.3.236

Balleine, B. W., & Killcross, S. (2006). Parallel incentive processing: an integrated view of amygdala function. *Trends in neurosciences*, *29*(5), 272-279.
https://doi.org/10.1016/j.tins.2006.03.002

Barrett, L. F. (2017). The theory of constructed emotion: an active inference account of interoception and categorization. *Social cognitive and affective neuroscience*, *12*(1), 1-23. https://doi.org/10.1093/scan/nsw154

Barrett, L. F., & Simmons, W. K. (2015). Interoceptive predictions in the brain. *Nature reviews neuroscience, 16(7),* 419-429. https://doi.org/10.1038/nrn3950

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin, 117*(3), 497–529. https://doi.org/10.1037/0033-2909.117.3.497

Becker, S., Bräscher, A. K., Bannister, S., Bensafi, M., Calma-Birling, D., Chan, R. C., ... & Wang, Y. (2019). The role of hedonics in the Human Affectome. *Neuroscience & Biobehavioral Reviews*, *102*, 221-241. https://doi.org/10.1016/j.neubiorev.2019.05.003

Berridge, K. C. (1996). Food reward: brain substrates of wanting and liking. *Neuroscience & Biobehavioral Reviews, 20(1),* 1-25. https://doi.org/10.1016/0149-7634(95)00033-B

Berridge, K. C. (2004). Motivation concepts in behavioral neuroscience. *Physiology & behavior, 81(2),* 179-209. https://doi.org/10.1016/j.physbeh.2004.02.004

Berridge, K. C. (2007). The debate over dopamine's role in reward: the case for incentive salience. *Psychopharmacology*, *191*(3), 391-431. https://doi.org/10.1007/s00213-006-0578-x

Berridge, K. C. (2018). Evolving concepts of emotion and motivation. Frontiers in Psychology, 9, 1647. https://doi.org/10.3389/fpsyg.2018.01647

Berridge, K. C., & Aldridge, J. W. (2009). Decision utility, incentive salience, and cue-triggered "wanting". *Oxford series in social cognition and social neuroscience, 2009,* 509.

Berridge, K. C., & Kringelbach, M. L. (2015). Pleasure systems in the brain. *Neuron*, *86*(3), 646-664. https://doi.org/10.1016/j.neuron.2015.02.018

Bindra, D. (1974). A motivational view of learning, performance, and behavior modification. *Psychological review, 81(3),* 199.
https://doi.org/10.1037/h0036330

Bosulu, J., Allaire, M. A., Tremblay-Grénier, L., Luo, Y., Eickhoff, S., & Hétu, S. (2022). 'Wanting'versus' Needing'related value: an fMRI meta-analysis. *Brain and Behavior*.
https://doi.org/10.1002/brb3.2713

Bouton, M. E. (2016). *Learning and behavior: A contemporary synthesis* (2nd ed.). Sinauer Associates.

Cabanac, M. (2017). Pleasure and joy, and their role in human life. In *Creating the productive workplace* (pp. 73-82). Routledge.

Campbell, C. (1998). Consumption and the Rhetorics of Need and Want. *Journal of design history, 11(3),* 235-246.
https://doi.org/10.1093/jdh/11.3.235

Chen, J., Papies, E. K., & Barsalou, L. W. (2016). A core eating network and its modulations underlie diverse eating phenomena. *Brain and cognition, 110*, 20-42.
https://doi.org/10.1016/j.bandc.2016.04.004

Craig, A. D. (2003). Interoception: the sense of the physiological condition of the body. *Current opinion in neurobiology, 13(4),* 500-505.
https://doi.org/10.1016/S0959-4388(03)00090-4

Demekas, D., Parr, T., & Friston, K. J. (2020). An investigation of the free energy principle for emotion recognition. *Frontiers in Computational Neuroscience*, *14*, 30.
https://doi.org/10.3389/fncom.2020.00030

Dickinson, A., & Balleine, B. (1994). Motivational control of goal-directed action. *Animal Learning & Behavior, 22(1),* 1-18.
https://doi.org/10.3758/BF03199951

Doya, K. (2000). Metalearning, neuromodulation, and emotion. In *Proc. Conf. Affect. Minds* (Vol. 46, p. 47).

FitzGerald, T. H., Dolan, R. J., & Friston, K. (2015). Dopamine, reward learning, and active inference. *Frontiers in computational neuroscience*, 136.
https://doi.org/10.3389/fncom.2015.00136

Friston, K., Kilner, J., & Harrison, L. (2006). A free energy principle for the brain. *Journal of physiology-Paris*, *100*(1-3), 70-87.
https://doi.org/10.1016/j.jphysparis.2006.10.001

Friston, K. (2010). The free-energy principle: a unified brain theory?. *Nature reviews neuroscience*, *11*(2), 127-138.
https://doi.org/10.1038/nrn2787

Friston, K., & Ao, P. (2012). Free energy, value, and attractors. *Computational and mathematical methods in medicine*, *2012*.
https://doi.org/10.1155/2012/937860

Haber, S. N., & Knutson, B. (2010). The reward circuit: linking primate anatomy and human imaging. *Neuropsychopharmacology*, *35*(1), 4-26.
https://doi.org/10.1038/npp.2009.129

Hamid, A. A., Pettibone, J. R., Mabrouk, O. S., Hetrick, V. L., Schmidt, R., Vander Weele, C. M., ... & Berke, J. D. (2016). Mesolimbic dopamine signals the value of work. *Nature neuroscience, 19(1),* 117-126.
https://doi.org/10.1038/nn.4173

Holmes, J. (2022). Friston's free energy principle: new life for psychoanalysis?. *BJPsych Bulletin*, *46*(3), 164-168.
https://doi:10.1192/bjb.2021.6

Li, J., & Daw, N. D. (2011). Signals in human striatum are appropriate for policy update rather than value prediction. *Journal of Neuroscience*, *31*(14), 5504-5511.
https://doi.org/10.1523/JNEUROSCI.6316-10.2011

Liu, Z., Lin, R., & Luo, M. (2020). Reward contributions to serotonergic functions. *Annu. Rev. Neurosci*, *43*(1), 141-162.
https://doi.org/10.1146/annurev-neuro-093019-112252

Livneh, Y., Sugden, A. U., Madara, J. C., Essner, R. A., Flores, V. I., Sugden, L. A., ... & Andermann, M. L. (2020). Estimation of current and future physiological states in insular cortex. *Neuron, 105(6),* 1094-1111.
https://doi.org/10.1016/j.neuron.2019.12.027

Lex, A., & Hauber, W. (2008). Dopamine D1 and D2 receptors in the nucleus accumbens core and shell mediate Pavlovian-instrumental transfer. *Learning & Memory*, *15*(7), 483-491. https://doi.org/10.1101/lm.978708

Luo, M., Li, Y., & Zhong, W. (2016). Do dorsal raphe 5-HT neurons encode "beneficialness"?. *Neurobiology of learning and memory*, *135*, 40-49. https://doi.org/10.1016/j.nlm.2016.08.008

MacGregor, D. (1960). *The human side of enterprise* (Vol. 21, No. 166.1960). McGraw‐Hill: New York.

Mohebi, A., Pettibone, J. R., Hamid, A. A., Wong, J. M. T., Vinson, L. T., Patriarchi, T., ... & Berke, J. D. (2019). Dissociable dopamine dynamics for learning and motivation. *Nature*, *570*(7759), 65-70. https://doi.org/10.1038/s41586-019-1235-y

Panksepp, J. (2004). *Affective neuroscience: The foundations of human and animal emotions*. Oxford university press.

Parr, T., & Friston, K. J. (2018). The anatomy of inference: generative models and brain structure. *Frontiers in computational neuroscience*, 90. https://doi.org/10.3389/fncom.2018.00090

Parr, T., Pezzulo, G., & Friston, K. J. (2022). *Active inference: the free energy principle in mind, brain, and behavior*. MIT Press.

Passingham, R. E., & Wise, S. P. (2012). *The neurobiology of the prefrontal cortex: anatomy, evolution, and the origin of insight*. OUP Oxford.

Rice, M. E., Patel, J. C., & Cragg, S. J. (2011). Dopamine release in the basal ganglia. *Neuroscience*, *198*, 112-137. https://doi.org/10.1016/j.neuroscience.2011.08.066

Salamone, J. D., Correa, M., Yang, J. H., Rotolo, R., & Presby, R. (2018). Dopamine, effort-based choice, and behavioral economics: basic and translational research. *Frontiers in behavioral neuroscience, 12,* 52. https://doi.org/10.3389/fnbeh.2018.00052

Schultz, W. (1998). Predictive reward signal of dopamine neurons. *Journal of neurophysiology*, *80*(1), 1-27.

https://doi.org/10.1152/jn.1998.80.1.1

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science, 275(5306),* 1593-1599.
https://doi.org/10.1126/science.275.5306.1593

Schwartenbeck, P., FitzGerald, T. H., Mathys, C., Dolan, R., & Friston, K. (2015). The dopaminergic midbrain encodes the expected certainty about desired outcomes. *Cerebral cortex*, *25*(10), 3434-3445.
https://doi.org/10.1093/cercor/bhu159

Sterling, P. (2004). Principles of allostasis: Optimal design, predictive regulation, pathophysiology, and rational therapeutics. *Allostasis, homeostasis, and the costs of physiological adaptation*, *17*, 17-64.

Sterling, P., & Laughlin, S. (2015). *Principles of neural design*. MIT press.

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press.

Swanson, L. W. (2000). Cerebral hemisphere regulation of motivated behavior. *Brain research*, *886*(1-2), 113-164.

Toates, F. (1994). Comparing motivational systems—An incentive motivation perspective. In C. R. Legg & D. A. Booth (Eds.), *Appetite: Neural and behavioural bases* (pp. 305–327). Oxford University Press.
https://doi.org/10.1093/acprof:oso/9780198547877.003.0013

Wassum, K. M., Ostlund, S. B., Balleine, B. W., & Maidment, N. T. (2011). Differential dependence of Pavlovian incentive motivation and instrumental incentive learning processes on dopamine signaling. *Learning & memory*, *18*(7), 475-483.
https://doi.org/10.1101/lm.2229311

Zhang, J., Berridge, K. C., Tindell, A. J., Smith, K. S., & Aldridge, J. W. (2009). A neural computational model of incentive salience. *PLoS computational biology, 5(7*), e1000437.
https://doi.org/10.1371/journal.pcbi.1000437