

# Obesity and environmental uncertainty

Luis Nicolás Luarte Rodríguez

## Introduction

The main aim of this section is (1) to review empirical evidence that links environment uncertainty with food-seeking behavior, mainly by observing changes in food intake, (2) then, possible ways uncertainty is represented at a neural level, emphasizing on the reward system. (3) to link the previous points to obesity, reinforcement learning and delayed discounting models will be presented as a way to establish a framework to characterize obesity as related to environmental uncertainty and the consequent effects on food-seeking behavior and food intake. Finally, (4) a review on evidence pointing to the Hypocretin/Orexin system as a foraging controller is presented as the immediate mechanisms linking uncertainty and food intake. In the following sections, obesogenic environments are considered in light of the previous framework, and specific types and characteristics of such environments will be pointed.

## Food-seeking behavior and uncertainty

Uncertainty implies situations where there is partial or incomplete information. When considering agency in decision-making tasks, uncertainty, refers to the incomplete information about the outcome that a given decision will generate, and also incomplete information about the probability distribution governing such outcome, whereas ‘risk’ implies knowledge about such probability distribution (De Groot and Thurik 2018). A particular feature of risk is the inverted ‘U’ shape, where at outcome probability 0 or 1 is at minima, and at maxima when probability is 0.5, this derives from risk being measured as outcome variance (Preuschoff, Bossaerts, and Quartz 2006). If an agent is situated in a natural environment, it is unlikely to have complete information about the decision-outcome pairing, and thus is forced to generate estimates or more complex models about environment statistical properties.

When considering the case of an agent in a natural environment, one of the most relevant cases is that of foraging. Foraging implies an agent searching for resources in a partially known environment, with depleting resources, and the supposition that the agent seeks to maximize its resources in the most efficient possible way, while accounting for unknown resource distribution (Charnov 1976). Multiple theories on how an agent must decide to

optimally allocate time to each resource patch (Wajnberg, Fauvergue, and Pons 2000), optimal path (Hills, Kalff, and Wiener 2013; Humphries and Sims 2014) and evolutionary roots (Wosniack et al. 2017). However, how environmental variables, such as resources uncertainty, modulate agent decision making have been less reviewed. Intuitively, if an agent does not have perfect information about the environment to be able to predict if there is going to be food in the future, it should consider a proxy of this, such as current food availability, which integrated over a time span, indicates the level of uncertainty of food availability.

To consider the modulation of uncertainty in decision making, one needs to take into account that uncertainty in food availability is a direct product of food scarcity, as the probability of food encounters is reduced proportional to scarcity. Thus, is to be expected that uncertainty effects are in line with food scarcity signaling, that is, a expected reduction in energy expenditure in order to preserve the energetic balance. However empirical evidence is not so clear in this regard (Polo 2002), as both increases and reductions in body mass given increased levels of uncertainty in food availability (Fokidis et al. 2012). When controlling for total food intake in condition of fixed or variable food availability, studies with birds show a decrease in body mass, which is explained, in part, by increased locomotor activity (Fokidis et al. 2012), this results are discussed in terms of stress (augmented due to food variability). However they can also be interpreted of expected foraging behavior.

When an environment presents higher uncertainty on food availability, it could be expected that food-seeking bouts are to be increased, in order to compensate for reduced food encounters or, complementary to that, an increased hoarding-type behavior. In the face of uncertainty, specifically, regarding food resources, a typical behavior is to increase food-seeking bouts, and resulting hoarding-type behavior, allowing the consumption of extra calories. This has been considered as a mechanism to prevent possible starvation because uncertainty is used as a proxy of future food scarcity, so eating in excess could prevent starvation (Anselme and Güntürkün 2019). However, this hoarding-behavior can also be explained by directly estimating food availability in the environment. Food scarcity does not necessarily act as an indicator of uncertainty, as food resources, while fewer, can be found in predictable places or at periodical time intervals. Nevertheless, when previous regularities regarding feeding routine, such as feeders position is constantly changed, increased intake is observed relative to the unchanging environment (Forkman 1993). This shows that, while food scarcity could be a cause of hoarding behavior, environment properties are enough to trigger such behaviors. In the previous case, in addition to altering the environment, food reserves were also altered. However, its levels were more than sufficient to satisfy energetic demands. Nevertheless, the uncertainty effect on food intake holds even when food levels are equated through predictable and unpredictable settings (Cuthill 2000). In addition to increased intake, under uncertainty,

foraging patterns, under mathematical modelling, are changed towards the start of day to account for missed foraging bouts (Bednekoff and Houston 1994).

Up to this point, there are at least 3 points to consider regarding food uncertainty, (1) under food availability uncertainty, food-seeking behavior is increased (Robinson et al. 2014; Fokidis et al. 2012; Polo 2002). (2) when uncertainty is increased strategies to maintain energetic balance, such as hoarding (Anselme and Güntürkün 2019) or increasing body mass (Moiron, Mathot, and Dingemanse 2018; Cuthill 2000) emerge. (3) such strategies imply a trade-off between preventing starvation and increasing risk of predation (due to reduced mobility), and as such are suggestive of a dynamic balance (Macleod et al. 2005).

To address food-seeking behavior under uncertainty, one needs to consider which behavior is reflective of a food-seeking action. In typical experimental settings, such action is reflected by interaction with cues or apparatuses that are related to reward delivery, when the action is interacting with a conditioned stimuli is called sign-tracking, whereas if interaction is with food dispenser is called goal-tracking (Silva, Silva, and Pear 1992). More specifically, sign-tracking, refers to an approaching behavior towards previously conditioned stimuli and rewards. So, it implies a previous conditional-stimulus and unconditional stimulus pairing, and, afterwards, tracking of the signal that was previously associated with the reward (Flagel 2014).

When uncertainty is introduced at the stage of conditional and unconditional stimulus pairing, as the probability of reward delivery upon lever pressing, sign-tracking increases as the probabilities of reward delivery approaches 50%, and the amount of reward is more varied (Anselme, Robinson, and Berridge 2013). In this case, as the delivery of a given reward gives no information about following one (delivery is determined by a probability function, independent of animal action), it can be assumed that, under Shannon entropy formulation, entropy (which can be understood as a measure of uncertainty) reaches the peak at 50% probability, and, furthermore, it predicts that uniform distributions, with more outcomes, increase uncertainty. Both were the case in the previously presented experiment (assuming uncertainty drove signal-tracking) as 50% probability of delivering 2 or 0 pellets had lower signal-tracking than 50% probability of delivering 0 or 1, 2, or 3 pellets with equal probability (16.7% for 1, 2 or 3 pellets). This, again, points out that increased food-seeking related behavior increases upon increased uncertainty even when food availability is controlled. This effect has been replicated in studies with amphetamine sensitization, where uncertainty (on conditioned stimulus and unconditioned stimulus) and sensitization, independently, augmented sign-tracking behavior, however, the effect of both uncertainty and sensitization was not additive suggesting a ceiling effect (Robinson et al. 2015).

The increased, food-seeking related behavior magnification by uncertainty, has been found

with partial reinforcement procedures (Collins et al. 1983), with manipulation of food placement variability (Forkman 1993), variability on reward quality and delivery delay (Craft 2016), and in sequential probability tasks (Stagner and Zentall 2010). Implying a robust effect across multiple food-related uncertainty scenarios.

From the perspective of a foraging animal, food sources are distributed in a partially known space, where effort must be made to obtain such sources. Uncertainty, reveals the consistency of food sources in a given space, where more uncertainty determines more difficulty in obtaining food. However, the consistency of food sources must be sensed through a mechanism that updates its estimate in a trial by trial basis, because is safe to assume that an agent interested in sensing environment uncertainty does not possess complete information. A plausible mechanism is to sense uncertainty, indirectly, via the reward prediction error. The reward prediction error is simply

$$actual\ reward - expected\ reward$$

As the reward prediction error is thought to operate in environments where a particular action lead to a probable reward, this error is used to update the value of any given action, then, the value of such time step (which can be associated with a given action) is given by the discounted rewards from that point onwards up to the termination of the trial series (Sutton and Barto 2018)

$$expected\ reward = reward_{t+1} + \gamma reward_{t+2} + \gamma^2 reward_{t+3} + \dots + \gamma^k reward_T$$

Here the trial series is composed of  $T$  time steps with a discount factor  $\gamma, 0 \leq \gamma \leq 1$ . The discount factor is there to signal the typical preference for obtaining rewards now rather than latter, on and how big it is will depend on properties of both agent and environment (Glimcher 2011).

The formulation presented above is just a mathematical representation of several assumptions of how an agent can learn expected reward values in a finite, trial based, experiment and then calculate the prediction error at each time step, how this reward prediction error is used to update values will be presented latter on. However, the main idea is that over trials, as the expected value approximates the real one, the reward prediction error goes down, nevertheless, if rewards value change the error goes up reflecting this change (see figure 1).

The main neural circuitry supporting the computation of the reward prediction error is thought to be supported, mainly, by the dopamine system (Schultz 2016). As the reward prediction error was first derived from behavioral data, to assess the biological feasibility three components must exist (1) expectation encoding units; (2) reward encoding units and (3) a subtraction unit (Watabe-Uchida, Eshel, and Uchida 2017)

Reward prediction errors models predict both three cases (1) where the expected reward and current reward are equal (no prediction error); (2) expected reward is less than the current reward (negative error) or (3) expected reward is greater than current reward (positive error). Midbrain dopamine neurons have been found to encode positive error but not negative under reinforcement learning models (Bayer and Glimcher 2005). Around this point two main hypothesis have been formulated, the first, proposes that negative error are encoded via lowering the fire-rate compared to the baseline (Schultz, Dayan, and Montague 1997), whereas the second, proposes an opponency system between dopamine and serotonin systems (Daw, Kakade, and Dayan 2002). Dopamine neurons in the ventral tegmental area have been found to encode the future discounted rewards (Enomoto et al. 2011). This two lines of evidence points that dopamine is capable of encoding expectation, reward value and doing subtraction (perhaps including the serotonin system), showing a significant complexity of this system, which might exceed value-related computations (Takahashi et al. 2017).

Above the function of dopamine neurons in reward prediction error has been stated, more specifically, this function seems to be related to the phasic activations, whereas, more sustained activation is related to reward uncertainty (measured as reward variance, thus reaching its peak at a probability of 0.5) (Fiorillo 2003). Such uncertainty-related signal has also been found in the orbito frontal cortex, amygdala (Schultz et al. 2008) and medial frontal lobe (Huettel 2005). A plausible hypothesis to link the reward prediction error and uncertainty encoding, can state that, over time, reward prediction error signals are integrated into an uncertainty signal, as, over time, more error is to be expected under higher reward variability (see figure ??). However, evidence points towards independent signals of reward prediction error and uncertainty in the orbito frontal cortex (Rushworth and Behrens 2008). Nevertheless, at least at a computational level, the reward prediction error can be used to estimate the reward-related uncertainty (Soltani and Izquierdo 2019).

As seen previously the reward prediction error is linked with uncertainty, as the error is a function over uncertainty levels. The way in that these two are connected can be seen in a more contextualized fashion. The main purpose of reward prediction error is to allow an agent to assess the value of actions in a certain environment, while measuring how stable is such environment. Thus, reward prediction error assists in learning, and in such context it serves the purpose of regulating the learning rate. Learning rate is parameter that defines how, a reward prediction error (or any measure of error), should affect posterior decisions, intuitively, when starting in a given task, higher learning rates are to be expected so learning happens at faster rate, however, as the task is properly learned learning rates should go down, so to not be influenced by random fluctuations (Even-Dar and Mansour 2001). Mathematical models, based on previously presented dopamine research, have proposed that learning rates are to be updated via the covariance between predictions (expected rewards) and prediction errors (Preuschoff and Bossaerts

2007), in the same vein, empirical experiments have shown that humans behave according to a, reward standard deviation dependent, scaling of reward prediction error (Diederen and Schultz 2015), so error should be less impactful when standard deviation is high. This proposed learning rate modifications are in line with the original model by Pearce and Hall (1980), which proposed that ‘surprise’ affected the learning rate, or viewed from the other side, as the pairing between unconditioned stimulus and conditioned stimulus became more predictable, the ‘associability’ decreased. Is important to note that other systems, aside from dopamine, are able to track environment uncertainty level, such as the endocrine system through the stress response, measured as subjective stress, pupil diameter and skin conductance (de Berker et al. 2016).

Up until now the discussion presented has focused on the calculation of reward value, typically, given a certain action. However, the computation of action-reward value is not directly linked to action choice. Picture a situation where the environment present high levels of uncertainty, and one action, until the present time step, have been associated with high rewards. If we were to choose just based on the maximum reward in such environment, we could miss potential better options, which true value, cannot be appropriately calculated because of variability. Such situation is more specifically defined by the exploration/exploitation dilemma, which posits that an agent, in order to obtain rewards, must ‘exploit’ current knowledge. However, it also must ‘explore’ to determine the best option in the future (Sutton and Barto 2018). One of the findings is that the manipulation of dopamine levels modulates striatal representation of reward prediction errors, and subjects treated with L-dopa (a metabolic precursor of dopamine) chose, with more frequency, the option with greater reward compared to the placebo group and haloperidol (dopamine receptor antagonist) group (Pessiglione et al. 2006). Although the authors did not report the temperature parameter (the one that determines the balance between exploration and exploitation) of the model, given that in all three group the optimal option was learned, it can be interpreted that the increase in dopamine levels effectively induced a bias toward exploitation. Direct evidence on the effect of L-dopa in exploration/exploitation parameters, effectively shows that exploration is reduced, and this is associated with modulation of uncertainty signals in the insula and anterior cingulate cortex (Chakroun et al. 2020).

Reward prediction error also contains information about the divergence between the current policy the animal is following to obtain food (or reward), and the optimal (or desired one). Such a system, then, is able to modify decision-making policies based on the reward-prediction error (Pessiglione et al. 2006).

Summing these ideas up, food-seeking behavior can be conceptualized as a series of decision-making actions that occur in an environment with varying degrees of uncertainty, and where each feeding bout is evaluated by the reward prediction error. Then, and

averaging over a history of feeding bouts, environment uncertainty can be obtained via the spread of the rewards relative to the mean. Evidence of neural systems actually sustaining this point that increased activity in dopaminergic systems can reflect uncertainty (Fiorillo 2003), and computational models have pointed the possibility of a basal ganglia circuitry that encodes mean rewards and its spread (Mikhael and Bogacz 2016).

## **Environment uncertainty and food intake**

### **Uncertainty representations at the neural level**

If uncertainty can modulate food-seeking behavior in order to increase intake and better sustain energetic reserves, it is expected to have at least two functional instances (1) an uncertainty sensing unit and (2) a reward processing unit, which can relay information to homeostatic-related and decision-making loci, to integrate such information and determine the next action to take. To determine the neural substrates of such instances, environment-agent dynamics can be represented through Markov decision tasks. Such tasks consider a set of states with possible transitions between each one, and two functions (1) the one in charge of determining the state transition given the agent action and (2) an action-state-reward function which maps a reward to a given action-state tuple. In such tasks, uncertainty is derived from probability matrices assigned to either of the two functions. When state transition functions are manipulated, two scenarios can be created (1) a regular one, where action-state transitions are deterministic, and (2) a random one, where action-state can not be predicted. Considering the latter, ventral striatal neurons can encode the prediction of immediate rewards, whereas the dorsal striatum is related to the former (Tanaka et al. 2006). This can be explained in terms of immediate and long-term reward prediction, as state transitions are random, subjects can only reliably predict the following rewards, in turn, if state-transition dynamics are deterministic, the reward of a long series of actions can be predicted.

If an environment is stable, then state-action-reward mappings can be optimized to reduce reward-prediction error. In this way, when the mapping is optimized, reward-related circuitry should reduce its activity (Friston 2009). However, this mapping is always modulated by environment dynamics regarding uncertainty. An optimal mapping in a given environment state can increase the reward prediction error in the same environment if this is non-stationary. The anterior cingulate cortex (ACC) has been shown to increase its activation levels when predictability in the environment drops (Davis, Choi, and Benoit 2010), effectively signaling environment dynamics.

As previously stated, environment dynamics need to be taken into account in order to appropriately interpret obtained rewards. If I visit a restaurant and the food served is delicious, my rating of the restaurant should not be too hasty as this could be just good luck. However, if this has always been the case, giving a high rating would be the correct

choice. In term of rewards, uncertainty is high when a given rewards give no information about the ones to come, conversely, certainty is achieved when a given reward gives all information about the following one. Direct tracking of environment volatility has been found to be well represented in the ACC (Behrens et al. 2007), presumably by encoding some sort of learning rate that bias valuation of rewards more to the short-term if volatility is high, and to the long-term if volatility is lower. The competing hypothesis of ACC describes its function to a decision-difficulty sensing unit, or demand of control when overriding default action is more optimal (Shenhav, Cohen, and Botvinick 2016). However, it should be noted that (???) results were circumscribed to the time point where the outcome is observed, which corresponds to the proper timing to assign obtained reward influence to the following behavior.

When representing the uncertainty of a given environment, an agent must pair the value obtained with the action performed. For each action possible, the agent updates the value of the action-reward tuple based on the reward prediction error.

Temporal dynamics of action-reward pairing and reward prediction error are such that the former occurs first relative to the later. Such temporal difference is reasonable because the pairing should be represented when taken action, and the prediction error requires feedback in order to compare obtained versus expected rewards. Considering this, the action-reward pairing has been found to be correlated to activity at the putamen, whereas rewards-prediction error, to be represented in the caudate nucleus (Haruno and Kawato 2006). However, as the authors point, both structures are likely to be involved in a larger loop containing the ACC, which would make sense to integrate reward evaluation over states, actions, and environmental uncertainty, and optimally influence following behavior.

It can be inferred from the way action-reward pairing is stated that it corresponds to action selection based on a history of rewards, which are mediated by the reward prediction error. Inhibition of putamen activity has effectively shown a reduction in performance when the task requires the consideration of reward history to select correct actions (Muranishi et al. 2011). Signal encoding, however, seems to be more complex, as basal ganglia direct pathway encode rewards outcomes, and the indirect pathway represents the next-action selection (Nonomura et al. 2018). Together, this points to a multi-structure network that represents expected and obtained rewards as an error, which allows easing computational requirements as the current state needs only to be compared with the expectation, that encompasses all previous history of rewards. Moreover, this signal updates rewards given actions, while considering environment volatility and the proper weighting of immediate versus long-term rewards. Thus, allowing to optimize behavior even when environments are non-stationary and rapidly changing.



# Models explaining food intake in obesity

## Reinforcement learning models

Temporal-difference learning models state how agents can estimate reward values in uncertain environments. At each time-step, the agent computes the value of a given state considering: (1) the estimated value (randomly initiated at first), and (2) the temporal-difference error, which represents the distance between the estimate of state value and the actual reward obtained in such state.

$$V(S_t) \leftarrow V(S_t) + \alpha(\text{Temporal Difference Error}) \quad (1)$$

$V(S_t)$  denotes the estimated value at a given state, and  $\alpha$  is used to model the agent learning rate. Additional parameter  $\rho$  has been proposed to model sensitivity to reward (Huys et al. 2013; Kroemer and Small 2016), such that the temporal difference error accounts for the subjective value of obtained rewards.

$$\text{Temporal Difference Error} = \rho \times \text{Reward} - V(S_t) \quad (2)$$

Obese subjects had shown reduced dorsal striatum activity to food rewards, which has been interpreted as reduced pleasure for food. However, simulations under the previously presented model show another option. That is, obese subjects show heightened reward sensitivity but decreased learning rates, ending in a lowered state value estimation (Kroemer and Small 2016). Modeled learning rates measures had shown that this is the case in obese subjects. Moreover, it points that negative prediction errors (the equivalent of temporal difference error) were used to a lesser extent than lean subjects, whereas positive errors showed no differences (Mathar et al. 2017). This can be interpreted as a difficulty to update reward or state values when the estimated reward is higher than the actual reward, possibly reflecting a short-term reward estimation.

It should be noted that more recent neuroimaging evidence points in favor of a hyper-reactivity of rewards circuitry, instead of hypo-reactivity. However, conclusions obtained by the model still hold, as such, hyper-reactivity is accompanied by a bias towards immediate rewards (Stice and Burger 2019). In line with the reinforcement learning model presented, evidence from probabilistic learning paradigms in obese subjects shows a decreased impact of negatively valued choices on consequent behavioral adaptation (Kube et al. 2018). These seemingly opposing results can stem from, previously not considered, quadratic associations between BMI and reward sensitivity, where an inverted U-shape is observed as BMI increases (Horstmann, Fenske, and Hankir 2015). Taken together, this finding suggests that obesity overfeeding is not only reliant on increased reward sensitivity (more reward sensitivity is assumed to increase intake), but other parameters such as

learning rates can determine the overall valuation of the reward, biasing decision-making to immediate rewards, that paired with highly palatable food can lead to excess caloric intake.

More complex models can include reward sensitivity in addition to different palatability indices of food encounters, effectively modeling the course of an agent with reward heterogeneity. Additionally, agents in this model are allowed to learn the value of different rewards values through different environments, such as highly-palatability, low-palatability, and mixed. Later on, the effects of the starting environment can be assessed. Results suggest that starting in an abundance of highly-palatable food slows the learning of food reward values in the following environments (Hammond et al. 2012). While this model does not inform about the effect on weight or intake levels, it shows how agents react to initial conditions or, more generally, to non-stationary environments.

### **Delayed discounting models**

Although the factors determining obesity as an outcome are multiple (Ang et al. 2013), it is reasonable to assume that the more immediate cause is excess intake relative to energetic demands. Moreover, excess intake is determined in an instance to instance basis, where a decision considering short and long-term benefits/risks must be made. With this in consideration, one can assume that obesity, in part, is caused by sub-optimal short/long-term benefit/risk assessments when making feeding decisions. If this was the case, as previously noted, areas that are related to computing options value in the short/long term, such as the ACC, should be in some way impaired.

Delayed discounting refers to the depreciation of a certain reward as a function of the time required to obtain it (da Matta, Gonçalves, and Bizarro 2012). As such, it provides measures of how reward-related systems bias decision to the short or long term. Obese subjects show a robust tendency to steeply discount future rewards (Amlung et al. 2016), thus, favoring short-term rewards.

Furthermore, ACC, among other structures, shows relative atrophy in obese subjects (Wang et al. 2017; Raji et al. 2009), suggesting an impairment of the previously mentioned functions. These findings can be interpreted as if impairment in environment uncertainty assessment results in a preference for short-term rewards. If this were the case, palatable food sensory cues, which trigger food-intake, would dominate over more long-term modulated decisions, such as healthy food intake (Higgs 2016).

Higher future rewards discounting paired with increased motivation to work for food, predict higher caloric intake (Rollins, Dearing, and Epstein 2010), and this effect seems to hold even for low energy-density food (Epstein et al. 2014). The rate of reward discounting, thus, informs about the predisposition to increased energetic intake, independent of possible food-property related effects. Similar effects have been found in children (Best et al. 2012),

but not in adult males (Smulders, Boswell, and Henderson 2019). Moreover, these effects seem to be directly related to body fat (Rasmussen, Lawyer, and Reilly 2010).

## **Orexin/Hypocretin system control of foraging**

Up to this point, the way reward-related systems interact with environmental uncertainty has been discussed. Several structures seem to be involved in integrating reward value in the face of environment volatility. Moreover, empirical findings of food-seeking behavior in predictable/unpredictable environments were pointed out. However, the direct mechanism that guides food-seeking behavior is lacking. One such system is the Orexin/Hypocretin (HO), which is part of the energetic homeostasis and feeding pathways (Toshinai et al. 2003), playing a large role in increasing food intake (Wolf 2009). However, a more broad and complex opioid system is thought to control food intake, which in turn is modulated by food preference, and has proven to be selective to certain macro-nutrients, such as fat (Taha 2010). More recent evidence has linked the activation of the hypothalamic HO system to an increase in short-term spatial memory, which is a function that supports exploratory foraging behavior (Aitta-aho et al. 2016).

Moreover, orexin promotion of such foraging-related behavior has been postulated as one of its main functions (Barson 2020). Such function is relevant because foraging behavior evolved in a specific type of environment, where resources are sparse, clustered, and is a potential risk of predation, and developed relatively stable strategies to deal with such conditions (Wosniack et al. 2017). Thus, foraging behavior seeks to generate a strategy to maximize energetic intake in a partially known environment. However, if environment resources are non-depleting, it can lead to behaviors such as binge eating, finally resulting in excess caloric intake (Barson 2020).

To provide a connection between food-seeking behavior and uncertainty, evidence on the effects of increasing such uncertainty on the proximal effect of food-seeking behavior, that is, food intake is necessary. In that regard, it was pointed out that, possibly because of survival mechanisms, environment uncertainty increased food intake and reduced energetic spending. Then, the sufficient functions to support such findings were discussed, emphasizing related structures and functions associated with each one. Obesity was associated with sharp delayed-discounting and ACC atrophy, which points towards a sub-optimal pairing between reward value assignments, given environment uncertainty levels. Also, the OH system role in foraging was discussed as a proximal cause of overfeeding. Together, this suggests that food-seeking behavior evolved to provide optimal decision-making strategies in uncertain and scarce environments. However, (1) when environment energetic density is high, such strategies would result in overfeeding, and (2) obesity in itself can impair homeostatic regulation by altering structures related to uncertainty and reward value processing. Previous points predict that underlying foraging mechanisms, in

certain environments, can lead to obesity.

## Obesogenic environments

### Cafeteria diet and uncertainty

### The decision making problem in obesity

## Conclusions

## Figures

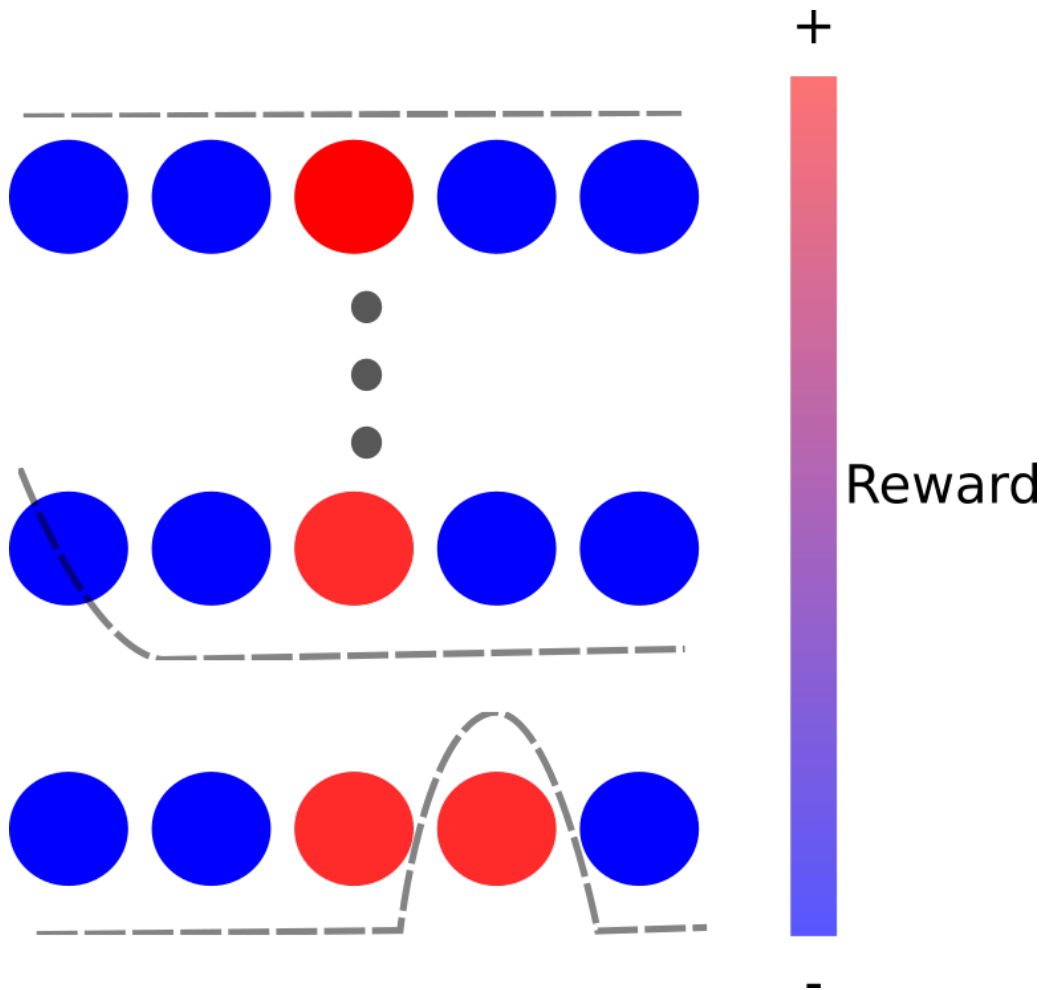


Figure 1: The figure illustrate how reward prediction error is diminished, as along a series of trials (circles denote trials, and three gray dots a series of trials) the reward value is learn and the error (doted gray lines) goes down. However when reward values are changed (last row) the reward prediction error goes up as this was not expected

## References

- Aitta-aho, Teemu, Elpiniki Pappa, Denis Burdakov, and John Apergis-Schoute. 2016. “Cellular Activation of Hypothalamic Hypocretin/Orexin Neurons Facilitates Short-Term Spatial Memory in Mice.” *Neurobiology of Learning and Memory* 136 (December): 183–88. <https://doi.org/10.1016/j.nlm.2016.10.005>.
- Amlung, M., T. Petker, J. Jackson, I. Balodis, and J. MacKillop. 2016. “Steep Discounting of Delayed Monetary and Food Rewards in Obesity: A Meta-Analysis.” *Psychological Medicine* 46 (11): 2423–34. <https://doi.org/10.1017/S0033291716000866>.
- Ang, Yeow Nyin, Bee Suan Wee, Bee Koon Poh, and Mohd Noor Ismail. 2013. “Multifactorial Influences of Childhood Obesity.” *Current Obesity Reports* 2 (1): 10–22. <https://doi.org/10.1007/s13679-012-0042-7>.
- Anselme, Patrick, and Onur Güntürkün. 2019. “How Foraging Works: Uncertainty Magnifies Food-Seeking Motivation.” *Behavioral and Brain Sciences* 42: e35. <https://doi.org/10.1017/S0140525X18000948>.
- Anselme, Patrick, Mike J. F. Robinson, and Kent C. Berridge. 2013. “Reward Uncertainty Enhances Incentive Salience Attribution as Sign-Tracking.” *Behavioural Brain Research* 238 (February): 53–61. <https://doi.org/10.1016/j.bbr.2012.10.006>.
- Barson, Jessica R. 2020. “Orexin/Hypocretin and Dysregulated Eating: Promotion of Foraging Behavior.” *Brain Research* 1731 (March): 145915. <https://doi.org/10.1016/j.brainres.2018.08.018>.
- Bayer, Hannah M., and Paul W. Glimcher. 2005. “Midbrain Dopamine Neurons Encode a Quantitative Reward Prediction Error Signal.” *Neuron* 47 (1): 129–41. <https://doi.org/10.1016/j.neuron.2005.05.020>.
- Bednekoff, Peter A., and Alasdair I. Houston. 1994. “Avian Daily Foraging Patterns: Effects of Digestive Constraints and Variability.” *Evolutionary Ecology* 8 (1): 36–52. <https://doi.org/10.1007/BF01237664>.
- Behrens, Timothy E J, Mark W Woolrich, Mark E Walton, and Matthew F S Rushworth. 2007. “Learning the Value of Information in an Uncertain World.” *Nature Neuroscience* 10 (9): 1214–21. <https://doi.org/10.1038/nn1954>.
- Best, John R., Kelly R. Theim, Dana M. Gredysa, Richard I. Stein, R. Robinson Welch, Brian E. Saelens, Michael G. Perri, Kenneth B. Schechtman, Leonard H. Epstein, and Denise E. Wilfley. 2012. “Behavioral Economic Predictors of Overweight Children’s Weight Loss.” *Journal of Consulting and Clinical Psychology* 80 (6): 1086–96. <https://doi.org/10.1037/a0029827>.
- Chakroun, Karima, David Mathar, Antonius Wiehler, Florian Ganzer, and Jan Peters.

2020. “Dopaminergic Modulation of the Exploration/Exploitation Trade-Off in Human Decision-Making.” *eLife* 9 (June): e51260. <https://doi.org/10.7554/eLife.51260>.
- Charnov, Eric L. 1976. “Optimal Foraging, the Marginal Value Theorem.” *Theoretical Population Biology* 9 (2): 129–36. [https://doi.org/10.1016/0040-5809\(76\)90040-X](https://doi.org/10.1016/0040-5809(76)90040-X).
- Collins, Louis, Douglas B. Young, Keith Davies, and John M. Pearce. 1983. “The Influence of Partial Reinforcement on Serial Autoshaping with Pigeons.” *The Quarterly Journal of Experimental Psychology Section B* 35 (4b): 275–90. <https://doi.org/10.1080/14640748308400893>.
- Craft, Baine B. 2016. “Risk-Sensitive Foraging: Changes in Choice Due to Reward Quality and Delay.” *Animal Behaviour* 111 (January): 41–47. <https://doi.org/10.1016/j.anbehav.2015.09.030>.
- Cuthill, I. C. 2000. “Body Mass Regulation in Response to Changes in Feeding Predictability and Overnight Energy Expenditure.” *Behavioral Ecology* 11 (2): 189–95. <https://doi.org/10.1093/beheco/11.2.189>.
- da Matta, Adriana, Fábio Leyser Gonçalves, and Lisiane Bizarro. 2012. “Delay Discounting: Concepts and Measures.” *Psychology & Neuroscience* 5 (2): 135–46. <https://doi.org/10.3922/j.psns.2012.2.03>.
- Davis, Jon F., Derrick L. Choi, and Stephen C. Benoit. 2010. “Insulin, Leptin and Reward.” *Trends in Endocrinology & Metabolism* 21 (2): 68–74. <https://doi.org/10.1016/j.tem.2009.08.004>.
- Daw, Nathaniel D, Sham Kakade, and Peter Dayan. 2002. “Opponent Interactions Between Serotonin and Dopamine.” *Neural Networks* 15 (4-6): 603–16. [https://doi.org/10.1016/S0893-6080\(02\)00052-7](https://doi.org/10.1016/S0893-6080(02)00052-7).
- de Berker, Archy O., Robb B. Rutledge, Christoph Mathys, Louise Marshall, Gemma F. Cross, Raymond J. Dolan, and Sven Bestmann. 2016. “Computations of Uncertainty Mediate Acute Stress Responses in Humans.” *Nature Communications* 7 (1): 10996. <https://doi.org/10.1038/ncomms10996>.
- De Groot, Kristel, and Roy Thurik. 2018. “Disentangling Risk and Uncertainty: When Risk-Taking Measures Are Not About Risk.” *Frontiers in Psychology* 9 (November): 2194. <https://doi.org/10.3389/fpsyg.2018.02194>.
- Diederen, Kelly M. J., and Wolfram Schultz. 2015. “Scaling Prediction Errors to Reward Variability Benefits Error-Driven Learning in Humans.” *Journal of Neurophysiology* 114 (3): 1628–40. <https://doi.org/10.1152/jn.00483.2015>.
- Enomoto, K., N. Matsumoto, S. Nakai, T. Satoh, T. K. Sato, Y. Ueda, H. Inokawa, M. Haruno, and M. Kimura. 2011. “Dopamine Neurons Learn to Encode the Long-Term

- Value of Multiple Future Rewards.” *Proceedings of the National Academy of Sciences* 108 (37): 15462–7. <https://doi.org/10.1073/pnas.1014457108>.
- Epstein, Leonard H., Noelle Jankowiak, Kelly D. Fletcher, Katelyn A. Carr, Chantal Nederkoorn, Hollie A. Raynor, and Eric Finkelstein. 2014. “Women Who Are Motivated to Eat and Discount the Future Are More Obese: BMI and Reinforcement Pathology.” *Obesity* 22 (6): 1394–9. <https://doi.org/10.1002/oby.20661>.
- Even-Dar, Eyal, and Yishay Mansour. 2001. “Learning Rates for Q-Learning.” In *Computational Learning Theory*, edited by David Helmbold and Bob Williamson, 2111:589–604. Berlin, Heidelberg: Springer Berlin Heidelberg. [http://link.springer.com/10.1007/3-540-44581-1\\_39](http://link.springer.com/10.1007/3-540-44581-1_39).
- Fiorillo, C. D. 2003. “Discrete Coding of Reward Probability and Uncertainty by Dopamine Neurons.” *Science* 299 (5614): 1898–1902. <https://doi.org/10.1126/science.1077349>.
- Flagel, Shelly B. 2014. “Sign-Tracking.” In *Encyclopedia of Psychopharmacology*, edited by Ian P. Stolerman and Lawrence H. Price, 1–7. Berlin, Heidelberg: Springer Berlin Heidelberg. [http://link.springer.com/10.1007/978-3-642-27772-6\\_7020-1](http://link.springer.com/10.1007/978-3-642-27772-6_7020-1).
- Fokidis, H. B., M. B. des Roziers, R. Sparr, C. Rogowski, K. Sweazea, and P. Devishe. 2012. “Unpredictable Food Availability Induces Metabolic and Hormonal Changes Independent of Food Intake in a Sedentary Songbird.” *Journal of Experimental Biology* 215 (16): 2920–30. <https://doi.org/10.1242/jeb.071043>.
- Forkman, B. A. 1993. “The Effect of Uncertainty on the Food Intake of the Mongolian Gerbil.” *Behaviour* 124 (3-4): 197–206. <https://doi.org/10.1163/156853993X00579>.
- Friston, Karl. 2009. “The Free-Energy Principle: A Rough Guide to the Brain?” *Trends in Cognitive Sciences* 13 (7): 293–301. <https://doi.org/10.1016/j.tics.2009.04.005>.
- Glimcher, P. W. 2011. “Understanding Dopamine and Reinforcement Learning: The Dopamine Reward Prediction Error Hypothesis.” *Proceedings of the National Academy of Sciences* 108 (Supplement\_3): 15647–54. <https://doi.org/10.1073/pnas.1014269108>.
- Hammond, Ross A., Joseph T. Ornstein, Lesley K. Fellows, Laurette Dubé, Robert Levitan, and Alain Dagher. 2012. “A Model of Food Reward Learning with Dynamic Reward Exposure.” *Frontiers in Computational Neuroscience* 6. <https://doi.org/10.3389/fncom.2012.00082>.
- Haruno, Masahiko, and Mitsuo Kawato. 2006. “Different Neural Correlates of Reward Expectation and Reward Expectation Error in the Putamen and Caudate Nucleus During Stimulus-Action-Reward Association Learning.” *Journal of Neurophysiology* 95 (2): 948–59. <https://doi.org/10.1152/jn.00382.2005>.
- Higgs, Suzanne. 2016. “Cognitive Processing of Food Rewards.” *Appetite* 104 (September): 10–17. <https://doi.org/10.1016/j.appet.2015.10.003>.

- Hills, Thomas T., Christopher Kalff, and Jan M. Wiener. 2013. "Adaptive Lévy Processes and Area-Restricted Search in Human Foraging." Edited by Yamir Moreno. *PLoS ONE* 8 (4): e60488. <https://doi.org/10.1371/journal.pone.0060488>.
- Horstmann, A., W. K. Fenske, and M. K. Hankir. 2015. "Argument for a Non-Linear Relationship Between Severity of Human Obesity and Dopaminergic Tone: Relationship Between Obesity and Dopaminergic Tone." *Obesity Reviews* 16 (10): 821–30. <https://doi.org/10.1111/obr.12303>.
- Huettel, S. A. 2005. "Decisions Under Uncertainty: Probabilistic Context Influences Activation of Prefrontal and Parietal Cortices." *Journal of Neuroscience* 25 (13): 3304–11. <https://doi.org/10.1523/JNEUROSCI.5070-04.2005>.
- Humphries, Nicolas E., and David W. Sims. 2014. "Optimal Foraging Strategies: Lévy Walks Balance Searching and Patch Exploitation Under a Very Broad Range of Conditions." *Journal of Theoretical Biology* 358 (October): 179–93. <https://doi.org/10.1016/j.jtbi.2014.05.032>.
- Huys, Quentin JM, Diego A Pizzagalli, Ryan Bogdan, and Peter Dayan. 2013. "Mapping Anhedonia onto Reinforcement Learning: A Behavioural Meta-Analysis." *Biology of Mood & Anxiety Disorders* 3 (1): 12. <https://doi.org/10.1186/2045-5380-3-12>.
- Kroemer, Nils B., and Dana M. Small. 2016. "Fuel Not Fun: Reinterpreting Attenuated Brain Responses to Reward in Obesity." *Physiology & Behavior* 162 (August): 37–45. <https://doi.org/10.1016/j.physbeh.2016.04.020>.
- Kube, Jana, David Mathar, Annette Horstmann, Sonja A. Kotz, Arno Villringer, and Jane Neumann. 2018. "Altered Monetary Loss Processing and Reinforcement-Based Learning in Individuals with Obesity." *Brain Imaging and Behavior* 12 (5): 1431–49. <https://doi.org/10.1007/s11682-017-9786-8>.
- Macleod, R., P. Barnett, J. A. Clark, and W. Cresswell. 2005. "Body Mass Change Strategies in Blackbirds *Turdus Merula*: The Starvation-Predation Risk Trade-Off." *Journal of Animal Ecology* 74 (2): 292–302. <https://doi.org/10.1111/j.1365-2656.2005.00923.x>.
- Mathar, David, Jane Neumann, Arno Villringer, and Annette Horstmann. 2017. "Failing to Learn from Negative Prediction Errors: Obesity Is Associated with Alterations in a Fundamental Neural Learning Mechanism." *Cortex* 95 (October): 222–37. <https://doi.org/10.1016/j.cortex.2017.08.022>.
- Mikhael, John G., and Rafal Bogacz. 2016. "Learning Reward Uncertainty in the Basal Ganglia." Edited by Kim T. Blackwell. *PLOS Computational Biology* 12 (9): e1005062. <https://doi.org/10.1371/journal.pcbi.1005062>.
- Moiron, Maria, Kimberley J. Mathot, and Niels J. Dingemanse. 2018. "To Eat and



- Not Be Eaten: Diurnal Mass Gain and Foraging Strategies in Wintering Great Tits.” *Proceedings of the Royal Society B: Biological Sciences* 285 (1874): 20172868. <https://doi.org/10.1098/rspb.2017.2868>.
- Muranishi, Manabu, Hitoshi Inokawa, Hiroshi Yamada, Yasumasa Ueda, Naoyuki Matsumoto, Masanori Nakagawa, and Minoru Kimura. 2011. “Inactivation of the Putamen Selectively Impairs Reward History-Based Action Selection.” *Experimental Brain Research* 209 (2): 235–46. <https://doi.org/10.1007/s00221-011-2545-y>.
- Nonomura, Satoshi, Kayo Nishizawa, Yutaka Sakai, Yasuo Kawaguchi, Shigeki Kato, Motokazu Uchigashima, Masahiko Watanabe, et al. 2018. “Monitoring and Updating of Action Selection for Goal-Directed Behavior Through the Striatal Direct and Indirect Pathways.” *Neuron* 99 (6): 1302–1314.e5. <https://doi.org/10.1016/j.neuron.2018.08.002>.
- Pearce, John M., and Geoffrey Hall. 1980. “A Model for Pavlovian Learning: Variations in the Effectiveness of Conditioned but Not of Unconditioned Stimuli.” *Psychological Review* 87 (6): 532–52. <https://doi.org/10.1037/0033-295X.87.6.532>.
- Pessiglione, Mathias, Ben Seymour, Guillaume Flandin, Raymond J. Dolan, and Chris D. Frith. 2006. “Dopamine-Dependent Prediction Errors Underpin Reward-Seeking Behaviour in Humans.” *Nature* 442 (7106): 1042–5. <https://doi.org/10.1038/nature05051>.
- Polo, V. 2002. “Daily Body Mass Regulation in Dominance-Structured Coal Tit (*Parus Ater*) Flocks in Response to Variable Food Access: A Laboratory Study.” *Behavioral Ecology* 13 (5): 696–704. <https://doi.org/10.1093/beheco/13.5.696>.
- Preuschoff, K., and P. Bossaerts. 2007. “Adding Prediction Risk to the Theory of Reward Learning.” *Annals of the New York Academy of Sciences* 1104 (1): 135–46. <https://doi.org/10.1196/annals.1390.005>.
- Preuschoff, Kerstin, Peter Bossaerts, and Steven R. Quartz. 2006. “Neural Differentiation of Expected Reward and Risk in Human Subcortical Structures.” *Neuron* 51 (3): 381–90. <https://doi.org/10.1016/j.neuron.2006.06.024>.
- Raji, Cyrus A., April J. Ho, Neelroop N. Parikshak, James T. Becker, Oscar L. Lopez, Lewis H. Kuller, Xue Hua, Alex D. Leow, Arthur W. Toga, and Paul M. Thompson. 2009. “Brain Structure and Obesity.” *Human Brain Mapping*, NA–NA. <https://doi.org/10.1002/hbm.20870>.
- Rasmussen, Erin B., Steven R. Lawyer, and William Reilly. 2010. “Percent Body Fat Is Related to Delay and Probability Discounting for Food in Humans.” *Behavioural Processes* 83 (1): 23–30. <https://doi.org/10.1016/j.beproc.2009.09.001>.
- Robinson, Mike J. F., Patrick Anselme, Adam M. Fischer, and Kent C. Berridge. 2014.

- “Initial Uncertainty in Pavlovian Reward Prediction Persistently Elevates Incentive Salience and Extends Sign-Tracking to Normally Unattractive Cues.” *Behavioural Brain Research* 266 (June): 119–30. <https://doi.org/10.1016/j.bbr.2014.03.004>.
- Robinson, Mike J. F., Patrick Anselme, Kristen Suchomel, and Kent C. Berridge. 2015. “Amphetamine-Induced Sensitization and Reward Uncertainty Similarly Enhance Incentive Salience for Conditioned Cues.” *Behavioral Neuroscience* 129 (4): 502–11. <https://doi.org/10.1037/bne0000064>.
- Rollins, Brandi Y., Kelly K. Dearing, and Leonard H. Epstein. 2010. “Delay Discounting Moderates the Effect of Food Reinforcement on Energy Intake Among Non-Obese Women.” *Appetite* 55 (3): 420–25. <https://doi.org/10.1016/j.appet.2010.07.014>.
- Rushworth, Matthew F S, and Timothy E J Behrens. 2008. “Choice, Uncertainty and Value in Prefrontal and Cingulate Cortex.” *Nature Neuroscience* 11 (4): 389–97. <https://doi.org/10.1038/nn2066>.
- Schultz, W., P. Dayan, and P. R. Montague. 1997. “A Neural Substrate of Prediction and Reward.” *Science* 275 (5306): 1593–9. <https://doi.org/10.1126/science.275.5306.1593>.
- Schultz, Wolfram. 2016. “Dopamine Reward Prediction Error Coding.” *Dialogues in Clinical Neuroscience* 18 (1): 23–32. <https://doi.org/27069377>.
- Schultz, Wolfram, Kerstin Preuschoff, Colin Camerer, Ming Hsu, Christopher D Fiorillo, Philippe N Tobler, and Peter Bossaerts. 2008. “Explicit Neural Signals Reflecting Reward Uncertainty.” *Philosophical Transactions of the Royal Society B: Biological Sciences* 363 (1511): 3801–11. <https://doi.org/10.1098/rstb.2008.0152>.
- Shenhav, Amitai, Jonathan D Cohen, and Matthew M Botvinick. 2016. “Dorsal Anterior Cingulate Cortex and the Value of Control.” *Nature Neuroscience* 19 (10): 1286–91. <https://doi.org/10.1038/nn.4384>.
- Silva, Francisco J., KathleenM. Silva, and Joseph J. Pear. 1992. “SIGN- VERSUS GOAL-TRACKING: EFFECTS OF CONDITIONED-STIMULUS-TO-UNCONDITIONED-STIMULUS DISTANCE.” *Journal of the Experimental Analysis of Behavior* 57 (1): 17–31. <https://doi.org/10.1901/jeab.1992.57-17>.
- Smulders, Tom V., Timothy Boswell, and Lindsay J. Henderson. 2019. “‘How Foraging Works’: Let’s Not Forget the Physiological Mechanisms of Energy Balance.” *Behavioral and Brain Sciences* 42: e51. <https://doi.org/10.1017/S0140525X1800198X>.
- Soltani, Alireza, and Alicia Izquierdo. 2019. “Adaptive Learning Under Expected and Unexpected Uncertainty.” *Nature Reviews Neuroscience* 20 (10): 635–44. <https://doi.org/10.1038/s41583-019-0180-y>.
- Stagner, Jessica P., and Thomas R. Zentall. 2010. “Suboptimal Choice Behavior by Pigeons.” *Psychonomic Bulletin & Review* 17 (3): 412–16. <https://doi.org/10.3758/>

- Stice, Eric, and Kyle Burger. 2019. "Neural Vulnerability Factors for Obesity." *Clinical Psychology Review* 68 (March): 38–53. <https://doi.org/10.1016/j.cpr.2018.12.002>.
- Sutton, Richard S., and Andrew G. Barto. 2018. *Reinforcement Learning: An Introduction*. Second edition. Adaptive Computation and Machine Learning Series. Cambridge, Massachusetts: The MIT Press.
- Taha, Sharif A. 2010. "Preference or Fat? Revisiting Opioid Effects on Food Intake." *Physiology & Behavior* 100 (5): 429–37. <https://doi.org/10.1016/j.physbeh.2010.02.027>.
- Takahashi, Yuji K., Hannah M. Batchelor, Bing Liu, Akash Khanna, Marisela Morales, and Geoffrey Schoenbaum. 2017. "Dopamine Neurons Respond to Errors in the Prediction of Sensory Features of Expected Rewards." *Neuron* 95 (6): 1395–1405.e3. <https://doi.org/10.1016/j.neuron.2017.08.025>.
- Tanaka, Saori C., Kazuyuki Samejima, Go Okada, Kazutaka Ueda, Yasumasa Okamoto, Shigeto Yamawaki, and Kenji Doya. 2006. "Brain Mechanism of Reward Prediction Under Predictable and Unpredictable Environmental Dynamics." *Neural Networks* 19 (8): 1233–41. <https://doi.org/10.1016/j.neunet.2006.05.039>.
- Toshinai, Koji, Yukari Date, Noboru Murakami, Mitsushi Shimada, Muhtashan S. Mondal, Takuya Shimbara, Jian-Lian Guan, et al. 2003. "Ghrelin-Induced Food Intake Is Mediated via the Orexin Pathway." *Endocrinology* 144 (4): 1506–12. <https://doi.org/10.1210/en.2002-220788>.
- Wajnberg, Eric, Xavier Fauvergue, and Odile Pons. 2000. "Patch Leaving Decision Rules and the Marginal Value Theorem: An Experimental Analysis and a Simulation Model." *Behavioral Ecology* 11 (6): 577–86. <https://doi.org/10.1093/beheco/11.6.577>.
- Wang, Haifeng, Baohong Wen, Jingliang Cheng, and Hongpeng Li. 2017. "Brain Structural Differences Between Normal and Obese Adults and Their Links with Lack of Perseverance, Negative Urgency, and Sensation Seeking." *Scientific Reports* 7 (1): 40595. <https://doi.org/10.1038/srep40595>.
- Watabe-Uchida, Mitsuko, Neir Eshel, and Naoshige Uchida. 2017. "Neural Circuitry of Reward Prediction Error." *Annual Review of Neuroscience* 40 (1): 373–94. <https://doi.org/10.1146/annurev-neuro-072116-031109>.
- Wolf, George. 2009. "Orexins: A Newly Discovered Family of Hypothalamic Regulators of Food Intake." *Nutrition Reviews* 56 (6): 172–73. <https://doi.org/10.1111/j.1753-4887.1998.tb06131.x>.
- Wosniack, Marina E., Marcos C. Santos, Ernesto P. Raposo, Gandhi M. Viswanathan, and Marcos G. E. da Luz. 2017. "The Evolutionary Origins of Lévy Walk Foraging."

Edited by Frederic Bartumeus. *PLOS Computational Biology* 13 (10): e1005774.  
<https://doi.org/10.1371/journal.pcbi.1005774>.