Obesity and environmental uncertainty

Luis Nicolás Luarte Rodríguez

Abstract

Obesity is a problem of great concern on a global scale. While it is considered a multifactorial disease, the proximal causes are mainly due to altered food consumption decisions. This review proposes a framework to merge the proximal causes with the environment-related distal causes of obesity. Drawing from foraging theory and reinforcement-learning models, considering decision-making as developed in its natural environment, and how this results in unwanted outcomes when palatable and varied food environments come into play. Additionally, a discussion of the neural basis of how the proximal and distal causes come into play. The proposed framework could help understand how the interplay between feeding-related behavior and the interaction with the environment could lead to obesity.

Introduction

Obesity prevalence increase and its economic cost (Withrow & Alter, 2011) present a significant concern ranging from society to individual health concerns. Despite its importance and the association between impaired decision making and obesity (C. Davis et al., 2004; Fitzpatrick et al., 2013), there is not a great wealth of literature including more refined models considering how decisions about food consumption could lead to obesity (Nettle et al., 2017). Thus, there exists a gap between the proximal causes of obesity and the decision-making processes that leads to it. Moreover, not having a model or theory of the underlying

decision-making processes further increases the gap between the underlying mechanisms controlling food consumption and how they interact with real world environments.

A first approximation to such models is to consider the underlying decision-making processes where they most likely evolved and the environmental settings that shaped them (Stephens, 2008). Foraging theory (Charnov, 1976) offers such possibility as it addresses how animals solve the problem of obtaining sufficient food to survive in environments that are not entirely known. As environment characteristics, such as food placement or predator threats are rarely known, foraging becomes a problem of searching in uncertain environments. While most animals can integrate sensory information and make decisions based on that, uncertainty makes such complex computations unreliable. Uncertainty, thus, bias organisms to simpler algorithms to pick the optimal option (Bartumeus et al., 2016) because the exploitation of environment structural characteristics allows the generation of decision rules that have minimal computational requirements (Fawcett et al., 2014).

The foraging problem can be briefly stated: animals must satisfy energetic demands while searching for food in an uncertain environment, balancing between overeating (becoming less mobile) and eating less but increasing the risk of death upon sudden food scarcity. To solve this, animals can exploit that food access uncertainty predicts food scarcity and have simple mechanisms that allow the computation of uncertainty, such as the reward prediction error (Colombo, 2014).

Understanding how the previously presented problem is solved and implemented offers a framework to explain how obesity is generated under particular environments. First, to develop such a framework, how food-seeking behavior relates to uncertainty and how animals deal with energetic balance is presented. Foraging theory offers an overview of the underlying processes of decision-making in natural environments. As uncertainty is a pivotal element of decision-making, how animals can assess uncertainty levels is discussed, drawing from prediction error hypothesis and reinforcement-learning theory. To connect uncertainty-related

computations with feeding-behavior, action selection algorithms are presented to state how uncertainty guides behavior. Then, neural implementations of previously discussed topics are proposed, mainly focusing on the prediction error. Finally, to tie up how uncertainty affects feeding behavior, obesogenic enironments' properties are reviewed in light of the decision models presented.

Food-seeking behavior and uncertainty

Uncertainty, as a broader concept, implies having limited or partial knowledge about something. Here, uncertainty will be understood as the knowledge of how a given action, inside a particular environment, relates to outcomes. For example, how does a lever press connects to food delivery? Pressing the lever always provides the food? Or only sometimes. Therefore, uncertainty represents the doubt about getting a food reward; when the lever is pressed, such doubt can be represented as a distribution over possible values (food or no-food) with an expected value, and a standard deviation to quantify the amount of uncertainty over the expected value. Note that standard deviation is just one among other metrics of uncertainty (Kerstin Preuschoff et al., 2006).

When considering agency in decision-making tasks, uncertainty, refers to the incomplete information about the outcome of a given decision, and also incomplete information about the probability distribution governing such outcome, whereas 'risk' implies knowledge about such probability distribution (De Groot & Thurik, 2018). A particular feature of risk is the inverted 'U' shape of its distribution. Where outcome probability 0 or 1 is at minima and maxima when the probability is 0.5, this derives from risk being measured as outcome variance between expected reward and actual reward (Kerstin Preuschoff et al., 2006). If an animal is situated in a natural environment, it is unlikely to have complete information about the action-outcome pairing, and thus, forced to generate estimates of such pairings, or to build more complex models about the environment.

Energetic balance in uncertain environments

Foraging is one of the most relevant cases of decision-making in natural environments. An animal foraging is equivalent to 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 an unknown resource distribution (Charnov, 1976). Multiple theories on how an agent must decide to optimally allocate time to each resource patch (Wajnberg et al., 2000), the optimal path to take (Hills et al., 2013; Humphries & Sims, 2014), and evolutionary roots of such strategies (Wosniack et al., 2017) have been largely studied in the literature. However, how environmental variables, such as resource uncertainty, modulates agent decision-making have been less reviewed. Likely due to a lack of integration over the fields of neuroscience, economics, and psychology (Rangel et al., 2008).

If an animal does not have perfect information about its environment, predicting if there will be food available in the near future, turns into a problem. To avoid such a situation, animals can consider how the current food availability fluctuates over time. If fluctuations are significant, it indicates high uncertainty over food availability; if small, food availability is certain. How animals could make use of such information? Contrary to experimental setups, uncertainty in food availability is probably a direct product of food scarcity in natural environments. The probability of having a food encounter is reduced proportional to resource levels, up to the point that levels are so low, there is certainty about resource depletion. Thus, it is expected that uncertainty effects are in line with food scarcity signaling, that is, an expected reduction in energy expenditure 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 have been observed (Fokidis et al., 2012), and this might be related to increased locomotor activity (Ferretti et al., 2019). When controlling for total food intake in fixed or variable food availability conditions, birds decrease their body mass, partly explained by increased locomotor activity

(Fokidis et al., 2012). Such an increase in locomotor activity can be a product of increased foraging bouts (Lohmus et al., 2006) or maintaining the number of foraging bouts but making them longer (Bordier et al., 2018).

Additionally, mathematical modeling of foraging behavior shows that an optimal strategy is to change foraging bouts to the start of the day, probably, to account for non-successful foraging bouts (Bednekoff & Houston, 1994), which has also been observed empirically (Ferretti et al., 2019). These results can be interpreted as stress being a signal of food availability and the modulation of foraging bouts to decreasing the efficiency of such bouts. Lower efficiency means that more bouts are to be made in order to sustain the energetic balance. While stress might signal uncertainty, the specific controller of foraging bouts, might be supported by the Locus Coeruleus activity, that disengages the animal from a given episode (Kane et al., 2017).

As previously stated, uncertainty could signal food scarcity. Thus, a compensatory strategy is necessary in order to survive when such signal arrives. Intuitively, if food is scarce more effort should be on finding food, that is, an increase in food-seeking behavior is expected. Complementary to putting more effort in searching for food, a hoarding-type behavior might aid to cope with diminished resources. Both strategies are considered mechanisms to prevent starvation in the near future, and are triggered by food-availability uncertainty, possibly because uncertainty is associated with food scarcity (Anselme & Güntürkün, 2019).

Food scarcity can generate multiple changes, such as the number of foraging spots visited, diet diversity, and others (T. R. Harris et al., 2010), which are similar to the ones already discussed. Nevertheless, there exist some differences between food scarcity and uncertainty. A scenario where the caloric density is low but constant exemplifies how an environment can have food scarcity but not uncertainty. Then, the question of what generates the previous behavioral changes arises. Evidence points out that these effects come from uncertainty rather than scarcity, as changes regarding feeding routine, such as feeder positions,

increase food intake compared to the constant environment (Forkman, 1993). Forkman (1993) study did not control for food-availability across conditions. However, in both conditions, food was enough to satisfy energetic demands. To control food-availability, equating total time available to get food has been used, and similar effects were reported when comparing predictable and unpredictable settings (Cuthill, 2000).

There are at least 3 points to consider regarding food uncertainty, (1) under food availability uncertainty, food-seeking behavior is increased (Fokidis et al., 2012; Polo, 2002; Mike J. F. Robinson et al., 2014). (2) when uncertainty is increased, strategies to maintain energetic balance, such as hoarding (Anselme & Güntürkün, 2019) or increasing body mass (Cuthill, 2000; Moiron et al., 2018) emerge. (3) such strategies imply a trade-off between preventing starvation and increasing the risk of predation (due to reduced mobility). As such, they suggest a dynamic balance between increasing body mass to prevent starvation and not doing so to retain mobility (Macleod et al., 2005).

Uncertainty increases food-seeking behavior

To more precisely address behaviors associated with uncertainty, what happens around the food-reward acquisition must be considered. That is, what are the behaviors that are typical when food availability is uncertain?. In experimental settings, such behaviors are considered as interactions with cues or apparatuses that are related to reward delivery. When the action is interacting with a conditioned stimulus (such as a lever or light), it is called sign-tracking, whereas if the interaction is with the food dispenser (or equivalent) is called goal-tracking (Silva et al., 1992). More specifically, sign-tracking refers to an approaching behavior towards previously conditioned stimuli and rewards. It implies a previous conditional-stimulus and unconditional-stimulus pairing and an afterward tracking of the previously associated signal with the reward (Flagel, 2014). This distinction is relevant because signal-tracking has been shown to respond robustly to uncertainty in food availability (Anselme et al., 2013).

When uncertainty is introduced at the stage of conditional and unconditional stimulus pairing, as the probability of obtaining a reward upon lever pressing, sign-tracking increases as the probabilities of reward delivery approaches 50%, and the amount of reward is more varied (Anselme et al., 2013). In this case, as the delivery of a given reward gives no information about the 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) (Namdari & Li, 2019) reaches its peak at 0.5 probability, and 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 0.5 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). Previous results, again, point out that increased food-seeking related behavior increases upon increased uncertainty even when food availability is controlled. Study replications 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 (Mike J. F. 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 & Zentall, 2010). Implying a robust effect across multiple food-related uncertainty scenarios.

Assessing and dealing with uncertainty

From the perspective of a foraging animal, food sources are distributed in a partially known space, where effort must be made to obtain such resources.

Uncertainty, reveals the consistency of food sources in a given space and time, where more uncertainty determines more difficulty in estimating current food availability. However, the consistency of food sources must be sensed through a mechanism that updates its estimates on a trial by trial basis. It 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 (Colombo, 2014). 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 leads to a probable reward, this error is used to update the value of any given action, then, the value at 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 & Barto, 2018)

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

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 later. How big it is will depend on both agent and environment (Glimcher, 2011).

The formulation presented above is just a mathematical representation of several assumptions on 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 later 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 such change (see figure 1).

The computation of reward predictions is dependent on dopamine system activity (Wolfram Schultz, 2016). As the reward prediction error was first derived from

behavioral data, to assess the biological feasibility of error computation, three components must exist (1) expectation encoding units; (2) reward encoding units, and (3) a subtraction unit (Watabe-Uchida et al., 2017)

Reward prediction errors models predict 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 errors but not negative under reinforcement learning models (Bayer & Glimcher, 2005). Around this point, two main hypotheses have been formulated. The first proposes that negative errors are encoded via lowering the fire-rate compared to the baseline (W. Schultz et al., 1997). Whereas the second proposes an opponency system between dopamine and serotonin systems (Daw et al., 2002). Furthermore, dopamine neurons in the ventral tegmental area have been found to encode the future discounted rewards (Enomoto et al., 2011). These 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).

The general 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 orbitofrontal cortex, amygdala (Wolfram 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 3). However, evidence points towards independent signals of reward prediction error and uncertainty in the orbito frontal cortex (Rushworth & Behrens, 2008). Nevertheless, at least

at a computational level, the reward prediction error can be used to estimate the reward-related uncertainty (Soltani & Izquierdo, 2019).

Prediction error is linked with uncertainty, as the error is expected to increase with higher uncertainty levels. Thus, prediction error allows an assessment of the value of actions and stability in an environment. To do this, only considering the absolute value of the error is not enough. While the prediction error provides a measure of the discrepancy between our prediction and actual values, the learning rate determines in what magnitude such error influences our estimates. If the environment is completely uncertain, past information is irrelevant, and recent information should hold more importance. Thus, learning rates should be high, so recent prediction errors influence the estimate to a greater measure. If the environment is certain, learning rates should be lower to represent historical information (Wu et al., 2017). Considering the development over the course of a task, when starting, higher learning rates are to be expected, so learning happens at a faster rate. However, as the task is properly learned, learning rates should go down not to be influenced by random fluctuations (Even-Dar & 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 (K. Preuschoff & Bossaerts, 2007). Model predictions are supported empirically, showing that humans behave according to a reward standard deviation-dependent scaling of reward prediction error (Diederen & Schultz, 2015), so the error should be less impactful when the standard deviation is high. The proposed learning rate modifications are in line with the original model by Pearce & 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. Learning rate increases as prediction error magnitude increases (Jepma et al., 2016). This allows for behavioral flexibility, as large errors signal environmental changes, and increasing the learning rates maximize

the importance of newer information. Is important to note that other systems, aside from dopamine, can 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), making uncertainty tracking via prediction error a possibility among others.

Changing the learning rate based on the reward prediction error reflects a constant tension an agent, faced with an uncertain environment, must face. How newly acquired information must be considered? A notion to answer this question is that of expected and unexpected uncertainty (Yu & Dayan, 2005). Expected uncertainty is the variability attributable to the stochastic nature of the reward. In contrast, unexpected uncertainty assumes that the agent creates beliefs about action-rewards associations, and incoming information breaks such beliefs (Payzan-LeNestour et al., 2013). Thus, unexpected uncertainty has been proposed as a top-down process that might be tracked by the Locus Coeruleus norepinephrine activity (Filipowicz et al., 2020; Payzan-LeNestour et al., 2013).

Moreover, such activity, measured as pupil diameter, has been found to track the learning rates (Nassar et al., 2012), representing the end product of assessing the environment in terms of expected and unexpected uncertainty. The intuition here is that when the statistical properties of an environment are changed, this should generate a signal of unexpected uncertainty similar to surprise, as proposed by Pearce & Hall (1980), which in turn represent that the current model of the environment must change to depict it accurately. The learning rate must then increase, giving more weight to more recent information and correctly updating the environment model (Faraji et al., 2017).

Until now, the discussion presented has focused on calculating reward value, given a specific action. However, the computation of action-reward value is not directly linked to action choice. Picture a situation where the environment presents high levels of uncertainty, and one action, until the present time step, has been associated with high rewards. Suppose we were to choose just based on the maximum reward in such an environment. In that case, we could miss

potentially better options, which true value cannot be appropriately calculated because of variability. Such a 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 & Barto, 2018). How exploration is controlled at a neural level is probably due to manipulating dopamine levels, which modulates striatal representation of reward prediction errors. 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 groups, 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).

Assessing uncertainty in the present and future

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 on 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, areas that are related to computing options value in short/long term, such as the ACC, should be in some way impaired.

Delayed discounting refers to the depreciation of a particular reward due to the

time required to obtain it (da Matta et al., 2012). As such, it provides measures of how reward-related systems bias decision to the short or long term. Obese subjects show a strong tendency to steeply discount future rewards (Amlung et al., 2016), thus, favoring short-term rewards.

Furthermore, among other structures, ACC shows relative atrophy in obese subjects (Raji et al., 2009; Wang et al., 2017), 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 et al., 2010), and this effect seems to hold even for low energy-density food (Epstein et al., 2014). Thus, the rate of reward discounting informs the predisposition to increase 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 et al., 2019). Moreover, these effects seem to be directly related to body fat (Rasmussen et al., 2010).

The bias towards immediate rewards

Previously, evidence on how uncertainty modulated feeding behavior, in terms of decisions between known and unknown options, and immediate and future valuation was presented. What follows aims to examine previous evidence in the case of overfeeding, specifically in obesity.

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.

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

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

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

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 & 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 & 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 (body mass index) and reward sensitivity, where an inverted U-shape is observed as BMI increases (Horstmann et al., 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. This, because, while palatable food definition is not standardized (Fazzino et al., 2019), it can be assumed that they, typically, consist of high caloric density. However, there might be additional effects of palatable foods in decreasing taste sensitivity related brain areas, which in turn, might favor further intake (Yokum & Stice, 2019).

Taking action in uncertain environments

Previously, the notion that there is something linking the estimated values and the actions taken was presented in terms of exploration/exploitation. How an agent decides, based on its estimated, to behave at any given time is called a 'policy,' and as such, it constitutes a mapping from estimates and actions (Sutton & Barto, 2018). As presented previously, the reward prediction error representation is able to guide the chosen policy of the agent (Pessiglione et al., 2006). Heuristics, which are strategies that rely heavily on exploiting environment statistical properties, have been proposed to be guiding decision-making in uncertain environments (Hafenbrädl et al., 2016). Some heuristics are thought to be an evolutionary derivative from uncertain environments (van den Berg & Wenseleers, 2018). However, what aspect of uncertainty is the one used to select the optimal policy is not clear (Gershman, 2018). Gershman (2018) explored the fit of two models to human behavioral data in two-armed bandit tasks and found signatures of both models in behavioral data, while a mixture of both models with more salient signatures represented the best fit. The first model fit corresponded to Upper-Confidence-Bound (Auer et al., 2002),

the intuition is that an agent should choose based on the times a certain action has been taken and the potential value of each action on the environment. The action selection is formally assigned as:

$$A_t := \operatorname*{argmax}_a \left[Q_t(a) + c\sqrt{\frac{ln_t}{N_t(a)}} \right]$$

Here $Q_t(a)$ represent the expected value of taking action a, in $c\sqrt{\frac{ln_t}{N_t(a)}}$ the denominator represents how many times action a has been chosen up to a certain point t, as time ln_t appears in the numerator, when action a is not chosen its upper-bound will increase, but decrease if such is continuously chosen. Note that Gershman (2018) used a modified version of this algorithm to reflect human decision stochasticity; nevertheless, this provides enough insight into how it considers uncertainty. The author found indirect support for this policy by considering that reaction times are faster when estimated rewards are more different (Tajima et al., 2016), and that reaction time decreased in proportion to increasing relative uncertainty, thus acting according to an uncertainty bonus as posed from the Upper-Confidence-Bound. The second model examined corresponded to Thompson Sampling, which builds reward priors on each option, this priors are beta-distributed with parameters α and β according to:

$$p(\theta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1}(1 - \theta)^{\beta-1},\, \theta \in [0, 1]$$

Where θ is the model expected reward, and Γ represents the Gamma function. To illustrate how the priors are updated, a win/loss reward environment can be considered. First, the agent will sample from each of the distribution and will choose the action associated with the distribution that gave the largest sample. Then, if the reward is a win (or '1') the α is updated as $\alpha = \alpha + 1$ and $\beta = \beta$, when the observed reward is a loss (or '0') $\beta = \beta + 1$ and $\alpha = \alpha$. Findings by Gershman (2018) noted that signatures corresponding to this model were that choice stochasticity (exploration actions) were proportional to the level of uncertainty in the model distributions.

Evidence for directed exploration, based on uncertainty levels, has been found in humans (Blanco & Sloutsky, 2019; Wilson et al., 2014). However, while directed exploration seems to be a robust strategy in humans, certain aspects emerge and vary throughout the life span (Somerville et al., 2017), pointing towards a complex and dynamic system. The main idea behind the previously presented models is twofold (1) uncertainty, in someway, guides the balance between exploration and exploitation, and (2) simple computations can sufficiently describe exploratory behavior under varying levels of uncertainty. Integrating over evidence of foraging under uncertainty and computational models presented, food-seeking behavior can be stated as a series of actions occurring in an uncertain environment, where each action (feeding bout) is evaluated in terms of the reward prediction error. Reward prediction error, however, not only informs about the value of the expected and current reward but also, by considering the history of such errors, calculates environmental uncertainty (Mikhael & Bogacz, 2016) (see figure 2) and represents it at a neural level (Fiorillo, 2003). Finally, action policies are intimately related to uncertainty, thus establishing a clear link between feeding-related behavior and environmental uncertainty.

Uncertainty representations at the neural level

Suppose uncertainty can modulate food-seeking behavior in order to increase intake and better sustain energetic reserves. In that case, it is expected to have at least to 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.

A plausible neuronal substrate underlying such functions is the striatum, as it is involved in decision-making actions such as action selection (Balleine et al., 2007), action-value representation (Samejima, 2005), and reward representation (Wake & Izuma, 2017). Markov decision tasks can be used to test behavior under certain or uncertain environments. Such a task comprises a set of states accessible by performing specific actions; however, the function that defines the mapping between actions and states can be random or deterministic. Finally, entering a given state provides a reward. When comparing learning of certain versus uncertain Markov decision tasks, the dorsal striatum seems to be more associated with the uncertain condition. In contrast, the dorsolateral prefrontal cortex showed greater activation under a certain condition (Tanaka et al., 2006). This can be explained in terms of immediate and long-term reward prediction. As state transitions are more uncertain, subjects can only reliably predict the following rewards. In turn, if state-transition dynamics are deterministic, the reward over a long series of actions can be predicted, making it useful to consider rewards in the long term. A similar relationship between uncertainty and immediate/long-term rewards has been noted in human consumers when exposed to features typically associated with environmental uncertainty, such as economic crisis, unemployment, among others (van den Berg & Wenseleers, 2018).

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 (J. F. Davis et al., 2010), effectively

signaling environment dynamics.

As previously stated, environment dynamics need to be taken into account in order to interpret obtained rewards appropriately. 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 terms 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 well represented in the ACC (Behrens et al., 2007), presumably by encoding some learning rate that bias valuation of rewards more to the short-term if volatility is high, and to the long-term is volatility is lower. ACC's competing hypothesis describes its function to a decision-difficulty sensing unit, or demand of control when overriding default action is more optimal (Shenhav et al., 2016). However, it should be noted that Behrens et al. (2007) 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 a given environment's uncertainty, an agent must pair the value obtained with the action performed. For each action possible, the agent updates the action-reward tuple's value 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 latter. Such a temporal difference is reasonable because the pairing should be represented when taken action, and the prediction error requires feedback to compare obtained versus expected rewards. Considering this, the action-reward pairing has been correlated to activity at the putamen, whereas rewards-prediction error is represented in the caudate nucleus (Haruno & Kawato, 2006). However, as the authors point, both structures are likely to be involved in a larger loop containing the ACC, making 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). However, signal encoding 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 representing expected and obtained rewards as an error, which allows easing computational requirements as the current state needs only to be compared with the expectation, which 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.

Orexin role in reward function

Orexin is a neuropeptide with two isoforms, orexin-A, and orexin-B, mainly associated with the regulation of arousal and feeding behaviors (Siegel, 2004). While orexin neurons are mainly localized in the lateral hypothalamus, they project widely throughout the brain (G. C. Harris & Aston-Jones, 2006; Nambu et al., 1999; Siegel, 2004). Lateral hypothalamus and orexin have been related to feeding and reward-related behavior G. C. Harris et al. (2005). Lateral hypothalamus regulates feed behavior (Anand & Brobeck, 1951; Margules & Olds, 1962; Sweeney & Yang, 2016), and does this by controlling orexin activity, where increased orexin activity increases feeding (Ardianto et al., 2016; Kotz et al., 2002). Furthermore, orexin-related arousal increase is modulated by extracellular glucose concentrations and neuroendocrine markers of energetic balance(Yamanaka et al., 2003), pointing to the inclusion of 'contextual' information to its function. On the other hand, the reward-related learning and valuation seem to be in

charge of the dopamine system and significantly sustained by the activity in the ventral tegmental area [D'Ardenne et al. (2008)'Ardenne_Etal_2008; Eshel et al. (2016)], that uses the reward prediction error to promote task learning (Keiflin et al., 2019). The ventral tegmental area is involved in more than the encoding of prediction errors, integrating more complex computation as the current beliefs about the environment (Gershman, 2017). Here, how 'contextual' information, such as environment uncertainty and animal current energetic balance, can affect how prediction errors and learning is assessed.

To the lateral hypothalamus to provide 'contextual' information to the ventral tegmental area would mean that (1) lateral hypothalamus encodes contextual information, (2) can relay this input to the ventral tegmental area, and (3) the input from the lateral hypothalamus modulates the excitability of dopamine neuron in the ventral tegmental area, thus altering the prediction error signal.

Lateral hypothalamus encodes contextual information   The signal carried by the lateral hypothalamus is complex as it encodes, among others, reward predictability and reward uncertainty (Noritake & Nakamura, 2019). This 'context' signal from the lateral hypothalamus to the ventral tegmental area, could explain why patients with narcolepsy-cataplexy (caused by the loss of orexin neurons of the lateral hypothalamus) show decreased performance in decision task with uncertainty, but not in normal decision tasks preserving reward sensitivity (Bayard et al., 2011). The latter is in line with the interpretation of lateral hypothalamus function as a 'predictive' homeostasis controller, feeding contextual information into the brain and signaling multiple control signals while having physiological state feedback (Burdakov, 2019). Such a function of the lateral hypothalamus is well supported by its anatomical disposition (Stefanie Geisler & Zahm, 2005).

Lateral hypothalamus relays information to the ventral tegmental area   Lateral hypothalamus projects a dense network of axons to the ventral tegmental area dopamine neurons, and this modulates reward-related behavior

via glutamatergic activity S. Geisler et al. (2007). Connectivity, with the same type of activity, exists between the lateral hypothalamus and lateral habenula, which may exert indirect control over the ventral tegmental area (Stamatakis et al., 2016). In addition to glutamatergic activity, those projections also contain a GABAergic element activated by leptin action (Leinninger et al., 2009). Among those innervations, a significant portion contains orexin neurons and is mainly limited to the ventral tegmental area (Fadel & Deutch, 2002). These projections have an important functional role as they are involved in compulsive sucrose seeking (Nieh et al., 2015), and more general behavioral activation throughout inhibitory GABAergics outputs from the lateral hypothalamus to the ventral tegmental area (Nieh et al., 2016).

The lateral hypothalamus orexin activity inputs into the ventral tegemental area via orexin receptor type 1 (Richardson & Aston-Jones, 2012), which allows an indirect modulation of the ventral tegemental area dopamine output to nucleus accumbens lateral shell, medial shell, and basolateral amygdala, effectively increasing activity in the first two, but not in the latter (Baimel et al., 2017). Thus, lateral hypothalamus orexin mediated activity might modulate reward-related behavior sustained by structures beyond the ventral tegmental area by influencing its output.

Previously how the hypothalamus encoded contextual information was considered. With recently presented information, the possibility of such contextual information to be carried into the ventral tegmental area is open. Furthermore, it points out that such information might modulate structures beyond the ventral tegmental area.

Lateral hypothalamus modulates the prediction error in the ventral tegmental area   Orexin activity increases palatable food intake when injected in the ventral tegmental area (Mattar et al., 2020; Terrill et al., 2016). Moreover, the orexin role in the lateral hypothalamus and ventral tegmental also modulated more general reward-related behavior (G. C. Harris et al., 2005; Mahmoudi et al.,

2020). Furthermore, activation of the lateral hypothalamus neurons projecting to the ventral tegmental area promotes reward-seeking behavior and displays a prediction error-like profile activity (Nieh et al., 2015). Lateral hypothalamus represents a prediction-error-like activity, and its connectivity to the ventral tegmental area prompts to think of direct modulation of the dopamine prediction error.

Lateral hypothalamus stores, previously learned, stimulus-reward pairings (Noritake & Nakamura, 2019; Sharpe et al., 2017), this can then inform about the expected value of a given action or stimulus in a previously experienced environment. If such information is not available, the prediction error magnitude should increase as the proper expectation is never adequately formed. An interesting case is when stored expected values are withdrawn at the time of prediction error computation but kept stored for the expected value update. Sharpe et al. (2017) observed this by optogenetically inactivating the lateral hypothalamus GABA neurons (which carry the expected value) terminals in the ventral tegmental area, only at cue presentation time in a typical pavlovian conditioning task, and found that mice spent more time in the food port upon positive conditioned stimulus. While counterintuitive, this case can be analogously understood as and increased learning rate, that is, having a larger prediction error (and thus increasing learning) can be obtained by increasing the learning rate or not considering the expected value in the computation, leading to a faster convergence to the estimated value. Because of the simple nature of the task, this procedure leads to increased performance, whoever faster convergence can lead to suboptimal decision making in more complex scenarios (Even-Dar & Mansour, 2001). This 'storage' of stimulus-value hypothalamic function is in line with evidence showing proportional orexin activity response to food or drug preference, and drug extinguished preference reinstatement (Aston-Jones et al., 2010), thus showing capabilities of encoding reward values and use them in a future time. This kind of signaling performed by the lateral hypothalamus adds up to a canonical role in feeding (Delgado & Anand, 1952; Jennings et al., 2013) into a more complex

connection to the reward systems, which, as previously mentioned, exceed the ventral tegmental area while sustaining the reward prediction error correlates, for example, in the nucleus accumbens (Werlen et al., 2020).

For orexin to modulate the reward function, specifically the prediction error, a series of conditions were considered, as the lateral hypothalamus's capabilities to encode contextual information and effectively input that into the ventral tegmental area. Moreover, it was considered that such input should alter the prediction error signal computed in the ventral tegmental area. All such capabilities are sustained by anatomical and functional evidence. As such, they posit the role of orexin in a pivotal role in the computation and update of rewards expected value. The primary importance of assessing the previously mentioned requisites is to add an essential part to the presented models: how environment information can be more directly taken into account at a neural level? Which adds up to the neural representations of uncertainty purely regarding reward-related computations.

The obesogenic environment

As discussed before, uncertainty is one of the most relevant features to assess in the environment because of its impact on survival. Because of its importance, animals and modern-day humans display a series of strategies to deal with environmental uncertainty, focusing on balancing insurance against scarcity (by increasing body fat deposits) and mobility (lower body weight improves mobility, saving against predation) (Brunstrom & Cheon, 2018). Sustaining insurance against scarcity strategy (increasing caloric intake) for prolonged periods, assuming sufficient caloric availability and equal energetic expenditure, leads to obesity, which is specifically defined by the percentage of body fat (Bhadoria et al., 2015). If such an insurance strategy were to respond to environmental uncertainty accurately, being introduced to a high caloric density environment should cease such a strategy. However, this often no the case.

Moreover, cafeteria-style diets, where the standard diet is supplemented with food rich in carbohydrates and fats, robustly generates obese animals (Hariri & Thibault, 2010; Leigh et al., 2019). Then, how come the 'wrong' strategy arises in food-abundant environments?.

To consider the previous question is useful to look in more detail the composition and effects of obesogenic environments. Drawing from drug-abuse literature results useful because of shared neural mechanisms, and phenotypes with obesity (Carlier et al., 2015; Filbey et al., 2012; García-García et al., 2014; Noori et al., 2016; Trinko et al., 2007). Considering such similarities, experimental evidence shows that prior exposure to uncertainty, measured as variance in food-reward variance in an operant task, generates enhanced nucleus accumbens dopamine activity. Increased nucleus accumbens activity is usually observed in psychostimulant exposure, paired with an increase in amphetamine self-administration compared with groups exposed to certain food delivery (Mascia et al., 2019). In natural situations, uncertainty usually translates to food insecurity, which refers to having food-availability affected by physical, economic, or other related factors (Barrett, 2010). Food insecurity has been related to increased BMI in wealthier countries, with a higher risk of obesity in women (Moradi et al., 2019), the increased association in women has been explained in terms of higher risk when having lower energetic reserves (Nettle et al., 2017). A simple explanation considers that wealthier countries have more ready access to food, so uncertainty-derived motivation to eat leads to effectively eating more (Nettle et al., 2017).

Considering the previous associations, an obesogenic environment might have two characteristics (1) is abundant in high caloric density foods or has an abundant food supply, and (2) it becomes obesogenic when previous experiences signal food uncertainty. How will this contextual information come into play when making decisions that will eventually lead to obesity?.

Cafeteria-type diet and high-fat diets induce obesity (Buyukdere et al., 2019; Rogers & Blundell, 1984). An intuitive interpretation is that both diets contain

energetically dense food, so every feeding-bout results in an increased caloric intake, which over-time would lead to obesity. However, a complimentary analysis of diet composition can tell how these diets differ in information and affect decision making differently. Cafeteria diets are typically more diverse, consisting of multiple palatable foods, whereas high-fat diets usually consist of fewer types of food. Considering this, the diversity in cafeteria diets might induce obesity more effectively than a fewer-food-item high-fat diet (Buyukdere et al., 2019). This effects has been observed across multiple scenarios including more variety on the food or substance of interest Dao et al. (2019). Divisive normalization can offer a model of how this case can affect decision making. The divisive normalization model assumes that values are encoded as firing rates (Anselme & Güntürkün, 2019), and that a given value is affected by other value representations in the following way:

$$\mu_i = K \frac{V_i}{\sigma_H + \Sigma_j w V_j}$$

This essentially shows that a given option value is normalized by (1) the number of other options and (2) the value of such options. μ_i represents the firing rate value encoding, while V_i is the currently encoded option, and V_j are the other options. Parameters K, w, and σ_H are used to generate realistic firing rates (Louie et al., 2013). In model simulations and monkey models of choice behavior, Louie et al. (2013) showed that introducing more options reduced the probability of choosing the highest value option. The previous results are interpreted in terms of the model, where increasing the number of options makes the values more similar. The previous model could be a potential mechanism to explain why increasing food variety increases intake. As food options are more varied, the expected value, and neural representations of every option becomes more homogeneous, making it harder to choose the highest value option (due to noise in value estimation). Thus, increasing the number of feeding-bouts needed to obtain the highest-valued or preferred option. It is important to note that mechanisms underlying intake increase due to variety are not known, and more

26

ecologically-leaned explanations are available (Michael G. Tordoff, 2002).

Conclusion

Uncertainty can be estimated by the ventral tegmental area and orexin activity in the lateral hypothalamus. The uncertainty level estimation is derived from prediction error and encoded via learning rate variation. As uncertainty rises, so does the learning rate, biasing towards short-term rewards. This way long-term health benefits are put in the background, compared to highly rewarding palatable food. Over time, such short-term bias could lead to obesity. Model for taking action suppose a second step after valuation, revised models further supported that uncertainty influences decision-making, biasing it towards exploratory behavior when environmental uncertainty is high. Increased exploratory behavior is consistent with presented studies of food-access uncertainty, showing increased locomotory activity.

When considering obesogenic environments, divisive normalization provided a model to explain a possible reason for the increased intake when food is more varied or when many options are available. Obesogenic environments, aside from being dense in energy-rich foods, make food value representations fuzzier, and, due to noise in value encoding, it is harder to pick the highest value option. This overlapping between options due to variety could generate increased intake by (1) increasing the number of feeding-bouts required to obtain the highest-valued option, or (2) noisier reward value estimation leading to either: (a) increased exploration to obtain a better estimate, or (b) increased prediction errors, due to worse value estimation, increasing the learning rate as if uncertainty was high, biasing decisions toward short-term rewards.
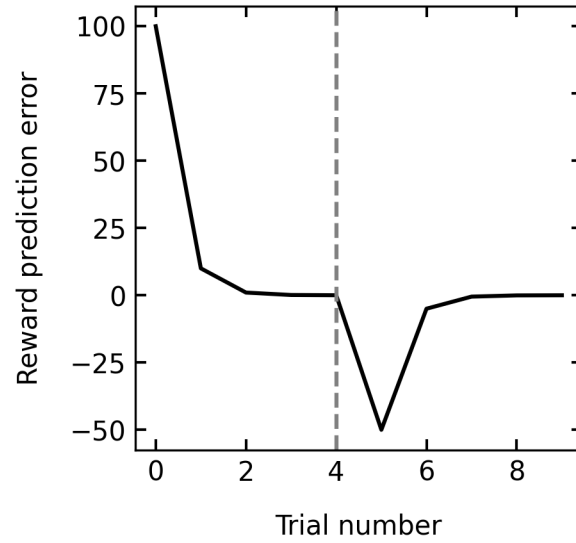
Figures

Figure 1: Figure shows a series of 10 trials, where, from trial 0-4, the true reward value is 100, and for the remaining trial its 50. A very basic agent was simulated to update its estimates based on the reward prediction error. Initial estimates were set at 0. Notice how during trial 0-4 reward prediction errors are positive and decrease to 0, because the reward obtained was, initially, greater than the estimate. In contrast, in trial 5, when reward changes to 50, the error becomes negative because the estimate was near 100
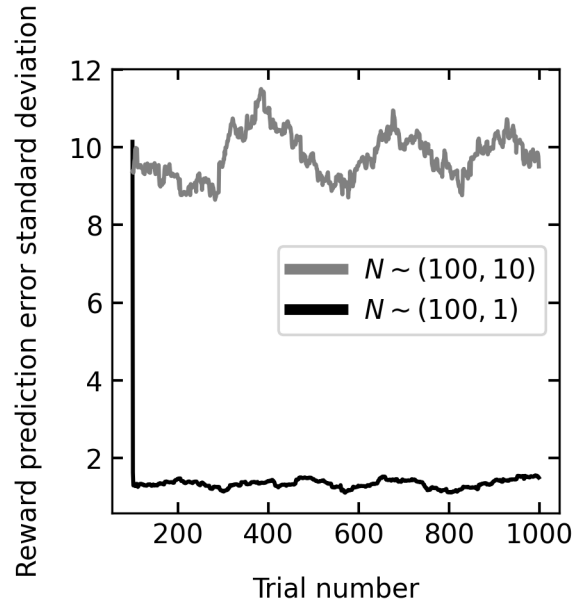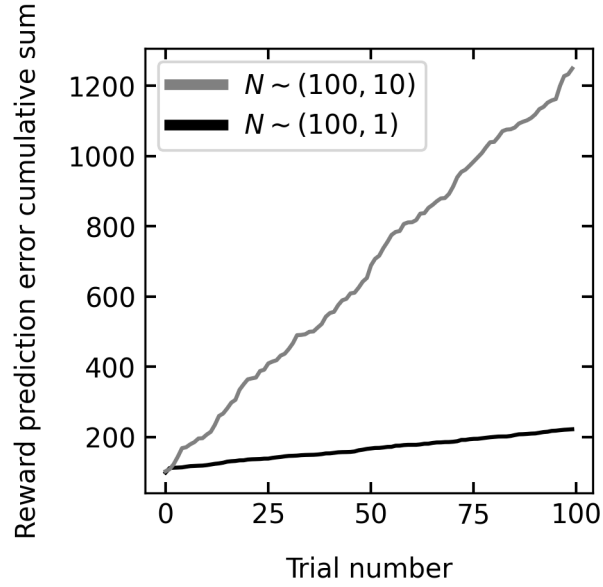
Figure 2: Simulated agent learning under two environments, (1) in gray, rewards are sampled from a normal distribution with mean = 100 and standard deviation = 10, (2) in black, mean = 100 and standard deviation = 1. Black and gray lines represent the reward prediction error rolling standard deviation over 100 trials. Notice how, over the trials, this 'signal' approximates the underlying uncertainty of the distribution (using the standard deviation as a measure of uncertainty).

Figure 3: Simulated agent learning under two environments, (1) in gray, rewards are sampled from a normal distribution with mean = 100 and standard deviation = 10, (2) in black, mean = 100 and standard deviation = 1. Agent in the environment with higher reward variability (gray), has a prediction error that varies proportionally to reward standard deviation, so its cumulative sum describes a steeper line than the agent drawing from lower standard deviation rewards. Note that under ideal situations, both agents will converge to the true reward value, however, in the more uncertain environment, this will happen at a slower pace.
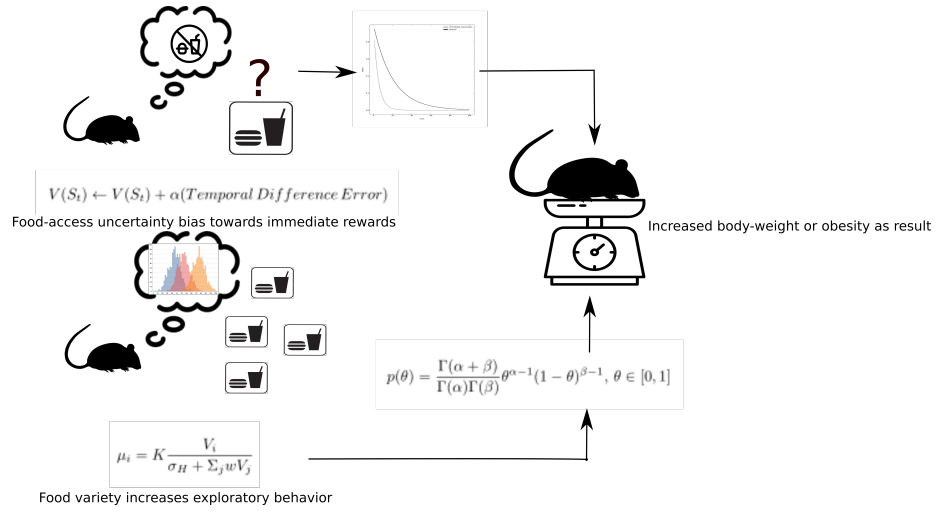
$$V(S_t) \leftarrow V(S_t) + \alpha(Temporal\ Difference\ Error)$$

Food-access uncertainty bias towards immediate rewards

Increased body-weight or obesity as result

$$p(\theta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1},\ \theta \in [0,1]$$

$$\mu_i = K\frac{V_i}{\sigma_H + \Sigma_j wV_j}$$

Food variety increases exploratory behavior

Figure 4: Graphical abstract

References

Amlung, M., Petker, T., Jackson, J., Balodis, I., & MacKillop, J. (2016). Steep discounting of delayed monetary and food rewards in obesity: a meta-analysis. *Psychological Medicine*, *46*(11), 2423–2434. https://doi.org/10.1017/S0033291716000866

Anand, B. K., & Brobeck, J. R. (1951). Localization of a "Feeding Center" in the Hypothalamus of the Rat. *Experimental Biology and Medicine*, *77*(2), 323–325. https://doi.org/10.3181/00379727-77-18766

Ang, Y. N., Wee, B. S., Poh, B. K., & Ismail, M. N. (2013). Multifactorial Influences of Childhood Obesity. *Current Obesity Reports*, *2*(1), 10–22. https://doi.org/10.1007/s13679-012-0042-7

Anselme, P., & Güntürkün, O. (2019). How foraging works: Uncertainty magnifies food-seeking motivation. *Behavioral and Brain Sciences*, *42*, e35. https://doi.org/10.1017/S0140525X18000948

Anselme, P., Robinson, M. J. F., & Berridge, K. C. (2013). Reward uncertainty enhances incentive salience attribution as sign-tracking. *Behavioural Brain Research*, *238*, 53–61. https://doi.org/10.1016/j.bbr.2012.10.006

Ardianto, C., Yonemochi, N., Yamamoto, S., Yang, L., Takenoya, F., Shioda, S., Nagase, H., Ikeda, H., & Kamei, J. (2016). Opioid systems in the lateral hypothalamus regulate feeding behavior through orexin and GABA neurons. *Neuroscience*, *320*, 183–193. https://doi.org/10.1016/j.neuroscience.2016.02.002

Aston-Jones, G., Smith, R. J., Sartor, G. C., Moorman, D. E., Massi, L., Tahsili-Fahadan, P., & Richardson, K. A. (2010). Lateral hypothalamic orexin/hypocretin neurons: A role in reward-seeking and addiction. *Brain Research*, *1314*, 74–90. https://doi.org/10.1016/j.brainres.2009.09.106

Auer, P., Cesa-Bianchi, N., & Fischer, P. (2002). Finite-time Analysis of the Multiarmed Bandit Problem. *Machine Learning*, *47*(2/3), 235–256. https:

//doi.org/10.1023/A:1013689704352

Baimel, C., Lau, B. K., Qiao, M., & Borgland, S. L. (2017). Projection-Target-Defined Effects of Orexin and Dynorphin on VTA Dopamine Neurons. *Cell Reports*, *18*(6), 1346–1355. https://doi.org/10.1016/j.celrep.2017.01.030

Balleine, B. W., Delgado, M. R., & Hikosaka, O. (2007). The Role of the Dorsal Striatum in Reward and Decision-Making. *Journal of Neuroscience*, *27*(31), 8161–8165. https://doi.org/10.1523/JNEUROSCI.1554-07.2007

Barrett, C. B. (2010). Measuring Food Insecurity. *Science*, *327*(5967), 825–828. https://doi.org/10.1126/science.1182768

Bartumeus, F., Campos, D., Ryu, W. S., Lloret-Cabot, R., Méndez, V., & Catalan, J. (2016). Foraging success under uncertainty: search tradeoffs and optimal space use. *Ecology Letters*, *19*(11), 1299–1313. https://doi.org/10.1111/ele.12660

Bayard, S., Abril, B., Yu, H., Scholz, S., Carlander, B., & Dauvilliers, Y. (2011). Decision Making in Narcolepsy with Cataplexy. *Sleep*, *34*(1), 99–104. https://doi.org/10.1093/sleep/34.1.99

Bayer, H. M., & Glimcher, P. W. (2005). Midbrain Dopamine Neurons Encode a Quantitative Reward Prediction Error Signal. *Neuron*, *47*(1), 129–141. https://doi.org/10.1016/j.neuron.2005.05.020

Bednekoff, P. A., & Houston, A. I. (1994). Avian daily foraging patterns: Effects of digestive constraints and variability. *Evolutionary Ecology*, *8*(1), 36–52. https://doi.org/10.1007/BF01237664

Behrens, T. E. J., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. S. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience*, *10*(9), 1214–1221. https://doi.org/10.1038/nn1954

Best, J. R., Theim, K. R., Gredysa, D. M., Stein, R. I., Welch, R. R., Saelens, B. E., Perri, M. G., Schechtman, K. B., Epstein, L. H., & Wilfley, D. E. (2012). Behavioral economic predictors of overweight children's weight loss.

Journal of Consulting and Clinical Psychology, *80*(6), 1086–1096. https://doi.org/10.1037/a0029827

Bhadoria, A., Sahoo, K., Sahoo, B., Choudhury, A., Sufi, N., & Kumar, R. (2015). Childhood obesity: Causes and consequences. *Journal of Family Medicine and Primary Care*, *4*(2), 187. https://doi.org/10.4103/2249-4863.154628

Blanco, N. J., & Sloutsky, V. (2019). *Systematic Exploration and Uncertainty Dominate Young Children's Choices*. PsyArXiv. https://osf.io/72sfx

Bordier, C., Klein, S., Le Conte, Y., Barron, A. B., & Alaux, C. (2018). Stress decreases pollen foraging performance in honeybees. *The Journal of Experimental Biology*, *221*(4), jeb171470. https://doi.org/10.1242/jeb.171470

Brunstrom, J. M., & Cheon, B. K. (2018). Do humans still forage in an obesogenic environment? Mechanisms and implications for weight maintenance. *Physiology & Behavior*, *193*, 261–267. https://doi.org/10.1016/j.physbeh.2018.02.038

Burdakov, D. (2019). Reactive and predictive homeostasis: Roles of orexin/hypocretin neurons. *Neuropharmacology*, *154*, 61–67. https://doi.org/10.1016/j.neuropharm.2018.10.024

Buyukdere, Y., Gulec, A., & Akyol, A. (2019). Cafeteria diet increased adiposity in comparison to high fat diet in young male rats. *PeerJ*, *7*, e6656. https://doi.org/10.7717/peerj.6656

Carlier, N., Marshe, V. S., Cmorejova, J., Davis, C., & Müller, D. J. (2015). Genetic Similarities between Compulsive Overeating and Addiction Phenotypes: A Case for "Food Addiction?" *Current Psychiatry Reports*, *17*(12), 96. https://doi.org/10.1007/s11920-015-0634-5

Chakroun, K., Mathar, D., Wiehler, A., Ganzer, F., & Peters, J. (2020). Dopaminergic modulation of the exploration/exploitation trade-off in human decision-making. *eLife*, *9*, e51260. https://doi.org/10.7554/eLife.51260

Charnov, E. L. (1976). Optimal foraging, the marginal value theorem. *Theoretical Population Biology*, *9*(2), 129–136. https://doi.org/10.1016/0040-5809(76)9

0040-X

Collins, L., Young, D. B., Davies, K., & Pearce, J. M. (1983). The Influence of Partial Reinforcement on Serial Autoshaping with Pigeons. *The Quarterly Journal of Experimental Psychology Section B*, *35*(4b), 275–290. https://doi.org/10.1080/14640748308400893

Colombo, M. (2014). Deep and beautiful. The reward prediction error hypothesis of dopamine. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *45*, 57–67. https://doi.org/10.1016/j.shpsc.2013.10.006

Craft, B. B. (2016). Risk-sensitive foraging: changes in choice due to reward quality and delay. *Animal Behaviour*, *111*, 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), 189–195. https://doi.org/10.1093/beheco/11.2.189

D'Ardenne, K., McClure, S. M., Nystrom, L. E., & Cohen, J. D. (2008). BOLD Responses Reflecting Dopaminergic Signals in the Human Ventral Tegmental Area. *Science*, *319*(5867), 1264–1267. https://doi.org/10.1126/science.1150605

da Matta, A., Gonçalves, F. L., & Bizarro, L. (2012). Delay discounting: Concepts and measures. *Psychology & Neuroscience*, *5*(2), 135–146. https://doi.org/10.3922/j.psns.2012.2.03

Dao, M. C., Messer, E., Conigliaro, T., Sakaida, K., Ouellette, A. F., Himaras, V., Thiron, S., & Roberts, S. B. (2019). Different and Unequal: A Qualitative Evaluation of Salient Factors Influencing Energy Intake in Adults with Overweight and Obesity. *Nutrients*, *11*(6), 1365. https://doi.org/10.3390/nu11061365

Davis, C., Levitan, R. D., Muglia, P., Bewell, C., & Kennedy, J. L. (2004).

Decision-Making Deficits and Overeating: A Risk Model for Obesity. *Obesity Research*, *12*(6), 929–935. https://doi.org/10.1038/oby.2004.113

Davis, J. F., Choi, D. L., & Benoit, S. C. (2010). Insulin, leptin and reward. *Trends in Endocrinology & Metabolism*, *21*(2), 68–74. https://doi.org/10.1016/j.tem.2009.08.004

Daw, N. D., Kakade, S., & Dayan, P. (2002). Opponent interactions between serotonin and dopamine. *Neural Networks*, *15*(4-6), 603–616. https://doi.org/10.1016/S0893-6080(02)00052-7

de Berker, A. O., Rutledge, R. B., Mathys, C., Marshall, L., Cross, G. F., Dolan, R. J., & Bestmann, S. (2016). Computations of uncertainty mediate acute stress responses in humans. *Nature Communications*, *7*(1), 10996. https://doi.org/10.1038/ncomms10996

De Groot, K., & Thurik, R. (2018). Disentangling Risk and Uncertainty: When Risk-Taking Measures Are Not About Risk. *Frontiers in Psychology*, *9*, 2194. https://doi.org/10.3389/fpsyg.2018.02194

Delgado, J. M. R., & Anand, B. K. (1952). Increase of Food Intake Induced by Electrical Stimulation of the Lateral Hypothalamus. *American Journal of Physiology-Legacy Content*, *172*(1), 162–168. https://doi.org/10.1152/ajplegacy.1952.172.1.162

Diederen, K. M. J., & Schultz, W. (2015). Scaling prediction errors to reward variability benefits error-driven learning in humans. *Journal of Neurophysiology*, *114*(3), 1628–1640. https://doi.org/10.1152/jn.00483.2015

DiLeone, R. J., Georgescu, D., & Nestler, E. J. (2003). Lateral hypothalamic neuropeptides in reward and drug addiction. *Life Sciences*, *73*(6), 759–768. https://doi.org/10.1016/S0024-3205(03)00408-9

Enomoto, K., Matsumoto, N., Nakai, S., Satoh, T., Sato, T. K., Ueda, Y., Inokawa, H., Haruno, M., & Kimura, M. (2011). Dopamine neurons learn to encode the long-term value of multiple future rewards. *Proceedings of the*

National Academy of Sciences, *108*(37), 15462–15467. https://doi.org/10.1 073/pnas.1014457108

Epstein, L. H., Jankowiak, N., Fletcher, K. D., Carr, K. A., Nederkoorn, C., Raynor, H. A., & Finkelstein, E. (2014). Women who are motivated to eat and discount the future are more obese: BMI and Reinforcement Pathology. *Obesity*, *22*(6), 1394–1399. https://doi.org/10.1002/oby.20661

Eshel, N., Tian, J., Bukwich, M., & Uchida, N. (2016). Dopamine neurons share common response function for reward prediction error. *Nature Neuroscience*, *19*(3), 479–486. https://doi.org/10.1038/nn.4239

Even-Dar, E., & Mansour, Y. (2001). Learning Rates for Q-Learning. In D. Helmbold & B. Williamson (Eds.), *Computational Learning Theory* (Vol. 2111, pp. 589–604). Springer Berlin Heidelberg. http://link.springer.com/10 .1007/3-540-44581-1_39

Fadel, J., & Deutch, A. Y. (2002). Anatomical substrates of orexin–dopamine interactions: lateral hypothalamic projections to the ventral tegmental area. *Neuroscience*, *111*(2), 379–387. https://doi.org/10.1016/S0306-4522(02)000 17-9

Faraji, M., Preuschoff, K., & Gerstner, W. (2017). Balancing New Against Old Information: The Role of Surprise in Learning. *arXiv:1606.05642 [cs, q-Bio, Stat]*. http://arxiv.org/abs/1606.05642

Fawcett, T. W., Fallenstein, B., Higginson, A. D., Houston, A. I., Mallpress, D. E. W., Trimmer, P. C., & McNamara, J. M. (2014). The evolution of decision rules in complex environments. *Trends in Cognitive Sciences*, *18*(3), 153–161. https://doi.org/10.1016/j.tics.2013.12.012

Fazzino, T. L., Rohde, K., & Sullivan, D. K. (2019). Hyper-Palatable Foods: Development of a Quantitative Definition and Application to the US Food System Database. *Obesity*, *27*(11), 1761–1768. https://doi.org/10.1002/oby. 22639

Ferretti, A., Maggini, I., Lupi, S., Cardinale, M., & Fusani, L. (2019). The amount of available food affects diurnal locomotor activity in migratory songbirds during stopover. *Scientific Reports*, *9*(1), 19027. https://doi.org/10.1038/s41598-019-55404-3

Filbey, F. M., Myers, U. S., & DeWitt, S. (2012). Reward circuit function in high BMI individuals with compulsive overeating: Similarities with addiction. *NeuroImage*, *63*(4), 1800–1806. https://doi.org/10.1016/j.neuroimage.2012.08.073

Filipowicz, A. L., Glaze, C. M., Kable, J. W., & Gold, J. I. (2020). Pupil diameter encodes the idiosyncratic, cognitive complexity of belief updating. *eLife*, *9*, e57872. https://doi.org/10.7554/eLife.57872

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

Fitzpatrick, S., Gilbert, S., & Serpell, L. (2013). Systematic Review: Are Overweight and Obese Individuals Impaired on Behavioural Tasks of Executive Functioning? *Neuropsychology Review*, *23*(2), 138–156. https://doi.org/10.1007/s11065-013-9224-7

Flagel, S. B. (2014). Sign-Tracking. In I. P. Stolerman & L. H. Price (Eds.), *Encyclopedia of Psychopharmacology* (pp. 1–7). Springer Berlin Heidelberg. http://link.springer.com/10.1007/978-3-642-27772-6_7020-1

Fokidis, H. B., des Roziers, M. B., Sparr, R., Rogowski, C., Sweazea, K., & Deviche, P. (2012). Unpredictable food availability induces metabolic and hormonal changes independent of food intake in a sedentary songbird. *Journal of Experimental Biology*, *215*(16), 2920–2930. 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, K. (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

García-García, I., Horstmann, A., Jurado, M. A., Garolera, M., Chaudhry, S. J., Margulies, D. S., Villringer, A., & Neumann, J. (2014). Reward processing in obesity, substance addiction and non-substance addiction: Obesity, addictions and reward. *Obesity Reviews*, *15*(11), 853–869. https://doi.org/10.1111/obr.12221

Geisler, S., Derst, C., Veh, R. W., & Zahm, D. S. (2007). Glutamatergic Afferents of the Ventral Tegmental Area in the Rat. *Journal of Neuroscience*, *27*(21), 5730–5743. https://doi.org/10.1523/JNEUROSCI.0012-07.2007

Geisler, Stefanie, & Zahm, D. S. (2005). Afferents of the ventral tegmental area in the rat-anatomical substratum for integrative functions. *The Journal of Comparative Neurology*, *490*(3), 270–294. https://doi.org/10.1002/cne.20668

Gershman, S. J. (2018). Deconstructing the human algorithms for exploration. *Cognition*, *173*, 34–42. https://doi.org/10.1016/j.cognition.2017.12.014

Gershman, S. J. (2017). Dopamine, Inference, and Uncertainty. *Neural Computation*, *29*(12), 3311–3326. https://doi.org/10.1162/neco_a_01023

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), 15647–15654. https://doi.org/10.1073/pnas.1014269108

Hafenbrädl, S., Waeger, D., Marewski, J. N., & Gigerenzer, G. (2016). Applied Decision Making With Fast-and-Frugal Heuristics. *Journal of Applied Research in Memory and Cognition*, *5*(2), 215–231. https://doi.org/10.1016/j.jarmac.2016.04.011

Hariri, N., & Thibault, L. (2010). High-fat diet-induced obesity in animal models.

Nutrition Research Reviews, *23*(2), 270–299. https://doi.org/10.1017/S09544
22410000168

Harris, G. C., & Aston-Jones, G. (2006). Arousal and reward: a dichotomy in orexin function. *Trends in Neurosciences*, *29*(10), 571–577. https://doi.org/10.1016/j.tins.2006.08.002

Harris, G. C., Wimmer, M., & Aston-Jones, G. (2005). A role for lateral hypothalamic orexin neurons in reward seeking. *Nature*, *437*(7058), 556–559. https://doi.org/10.1038/nature04071

Harris, T. R., Chapman, C. A., & Monfort, S. L. (2010). Small folivorous primate groups exhibit behavioral and physiological effects of food scarcity. *Behavioral Ecology*, *21*(1), 46–56. https://doi.org/10.1093/beheco/arp150

Haruno, M., & Kawato, M. (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), 948–959. https://doi.org/10.1152/jn.00382.2005

Higgs, S. (2016). Cognitive processing of food rewards. *Appetite*, *104*, 10–17. https://doi.org/10.1016/j.appet.2015.10.003

Hills, T. T., Kalff, C., & Wiener, J. M. (2013). Adaptive Lévy Processes and Area-Restricted Search in Human Foraging. *PLoS ONE*, *8*(4), e60488. https://doi.org/10.1371/journal.pone.0060488

Horstmann, A., Fenske, W. K., & Hankir, M. K. (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), 821–830. 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), 3304–3311. https://doi.org/10.1523/JNEUROSCI.5070-04.2005

Humphries, N. E., & Sims, D. W. (2014). Optimal foraging strategies: Lévy

walks balance searching and patch exploitation under a very broad range of conditions. *Journal of Theoretical Biology*, *358*, 179–193. https://doi.org/10.1016/j.jtbi.2014.05.032

Huys, Q. J., Pizzagalli, D. A., Bogdan, R., & Dayan, P. (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

Jennings, J. H., Rizzi, G., Stamatakis, A. M., Ung, R. L., & Stuber, G. D. (2013). The Inhibitory Circuit Architecture of the Lateral Hypothalamus Orchestrates Feeding. *Science*, *341*(6153), 1517–1521. https://doi.org/10.1126/science.1241812

Jepma, M., Murphy, P. R., Nassar, M. R., Rangel-Gomez, M., Meeter, M., & Nieuwenhuis, S. (2016). Catecholaminergic Regulation of Learning Rate in a Dynamic Environment. *PLOS Computational Biology*, *12*(10), e1005171. https://doi.org/10.1371/journal.pcbi.1005171

Kane, G. A., Vazey, E. M., Wilson, R. C., Shenhav, A., Daw, N. D., Aston-Jones, G., & Cohen, J. D. (2017). Increased locus coeruleus tonic activity causes disengagement from a patch-foraging task. *Cognitive, Affective, & Behavioral Neuroscience*, *17*(6), 1073–1083. https://doi.org/10.3758/s13415-017-0531-y

Keiflin, R., Pribut, H. J., Shah, N. B., & Janak, P. H. (2019). Ventral Tegmental Dopamine Neurons Participate in Reward Identity Predictions. *Current Biology*, *29*(1), 93–103.e3. https://doi.org/10.1016/j.cub.2018.11.050

Kempadoo, K. A., Tourino, C., Cho, S. L., Magnani, F., Leininger, G.-M., Stuber, G. D., Zhang, F., Myers, M. G., Deisseroth, K., de Lecea, L., & Bonci, A. (2013). Hypothalamic Neurotensin Projections Promote Reward by Enhancing Glutamate Transmission in the VTA. *Journal of Neuroscience*, *33*(18), 7618–7626. https://doi.org/10.1523/JNEUROSCI.2588-12.2013

Kotz, C. M., Teske, J. A., Levine, J. A., & Wang, C. (2002). Feeding and activity

induced by orexin A in the lateral hypothalamus in rats. *Regulatory Peptides*, *104*(1-3), 27–32. https://doi.org/10.1016/S0167-0115(01)00346-9

Kroemer, N. B., & Small, D. M. (2016). Fuel not fun: Reinterpreting attenuated brain responses to reward in obesity. *Physiology & Behavior*, *162*, 37–45. https://doi.org/10.1016/j.physbeh.2016.04.020

Kube, J., Mathar, D., Horstmann, A., Kotz, S. A., Villringer, A., & Neumann, J. (2018). Altered monetary loss processing and reinforcement-based learning in individuals with obesity. *Brain Imaging and Behavior*, *12*(5), 1431–1449. https://doi.org/10.1007/s11682-017-9786-8

Leigh, S.-J., Kendig, M. D., & Morris, M. J. (2019). Palatable Western-style Cafeteria Diet as a Reliable Method for Modeling Diet-induced Obesity in Rodents. *Journal of Visualized Experiments*, *153*, 60262. https://doi.org/10.3791/60262

Leinninger, G. M., Jo, Y.-H., Leshan, R. L., Louis, G. W., Yang, H., Barrera, J. G., Wilson, H., Opland, D. M., Faouzi, M. A., Gong, Y., Jones, J. C., Rhodes, C. J., Chua, S., Diano, S., Horvath, T. L., Seeley, R. J., Becker, J. B., Münzberg, H., & Myers, M. G. (2009). Leptin Acts via Leptin Receptor-Expressing Lateral Hypothalamic Neurons to Modulate the Mesolimbic Dopamine System and Suppress Feeding. *Cell Metabolism*, *10*(2), 89–98. https://doi.org/10.1016/j.cmet.2009.06.011

Lohmus, M., Sundstrom, L. F., & Moore, F. R. (2006). Non-invasive corticosterone treatment changes foraging intensity in red-eyed vireos Vireo olivaceus. *Journal of Avian Biology*, *37*(5), 523–526. https://doi.org/10.1111/j.0908-8857.2006.03733.x

Louie, K., Khaw, M. W., & Glimcher, P. W. (2013). Normalization is a general neural mechanism for context-dependent decision making. *Proceedings of the National Academy of Sciences*, *110*(15), 6139–6144. https://doi.org/10.1073/pnas.1217854110

Macleod, R., Barnett, P., Clark, J. A., & Cresswell, W. (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

Mahmoudi, M., Maleki-Roveshti, M., Karimi-Haghighi, S., & Haghparast, A. (2020). Chemical stimulation of the lateral hypothalamus induced seeking behaviors in rats: Involvement of orexin receptors in the ventral tegmental area. *European Journal of Pharmacology*, *886*, 173433. https://doi.org/10.1016/j.ejphar.2020.173433

Margules, D. L., & Olds, J. (1962). Identical "Feeding" and "Rewarding" Systems in the Lateral Hypothalamus of Rats. *Science*, *135*(3501), 374–375. https://doi.org/10.1126/science.135.3501.374

Mascia, P., Neugebauer, N. M., Brown, J., Bubula, N., Nesbitt, K. M., Kennedy, R. T., & Vezina, P. (2019). Exposure to conditions of uncertainty promotes the pursuit of amphetamine. *Neuropsychopharmacology*, *44*(2), 274–280. https://doi.org/10.1038/s41386-018-0099-4

Mathar, D., Neumann, J., Villringer, A., & Horstmann, A. (2017). Failing to learn from negative prediction errors: Obesity is associated with alterations in a fundamental neural learning mechanism. *Cortex*, *95*, 222–237. https://doi.org/10.1016/j.cortex.2017.08.022

Mattar, P., Uribe-Cerda, S., Pezoa, C., Guarnieri, T., Kotz, C. M., Teske, J. A., Morselli, E., & Perez-Leighton, C. (2020). Brain site-specific regulation of hedonic intake by orexin and DYN peptides: role of the PVN and obesity. *Nutritional Neuroscience*, 1–10. https://doi.org/10.1080/1028415X.2020.1840049

Mikhael, J. G., & Bogacz, R. (2016). Learning Reward Uncertainty in the Basal Ganglia. *PLOS Computational Biology*, *12*(9), e1005062. https://doi.org/10.1371/journal.pcbi.1005062

Moiron, M., Mathot, K. J., & Dingemanse, N. J. (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

Moradi, S., Mirzababaei, A., Dadfarma, A., Rezaei, S., Mohammadi, H., Jannat, B., & Mirzaei, K. (2019). Food insecurity and adult weight abnormality risk: a systematic review and meta-analysis. *European Journal of Nutrition*, *58*(1), 45–61. https://doi.org/10.1007/s00394-018-1819-6

Muranishi, M., Inokawa, H., Yamada, H., Ueda, Y., Matsumoto, N., Nakagawa, M., & Kimura, M. (2011). Inactivation of the putamen selectively impairs reward history-based action selection. *Experimental Brain Research*, *209*(2), 235–246. https://doi.org/10.1007/s00221-011-2545-y

Nambu, T., Sakurai, T., Mizukami, K., Hosoya, Y., Yanagisawa, M., & Goto, K. (1999). Distribution of orexin neurons in the adult rat brain1Published on the World Wide Web on 17 March 1999.1. *Brain Research*, *827*(1-2), 243–260. https://doi.org/10.1016/S0006-8993(99)01336-0

Namdari, A., & Li, Z. (Steven). (2019). A review of entropy measures for uncertainty quantification of stochastic processes. *Advances in Mechanical Engineering*, *11*(6), 168781401985735. https://doi.org/10.1177/1687814019857350

Nassar, M. R., Rumsey, K. M., Wilson, R. C., Parikh, K., Heasly, B., & Gold, J. I. (2012). Rational regulation of learning dynamics by pupil-linked arousal systems. *Nature Neuroscience*, *15*(7), 1040–1046. https://doi.org/10.1038/nn.3130

Nettle, D., Andrews, C., & Bateson, M. (2017). Food insecurity as a driver of obesity in humans: The insurance hypothesis. *Behavioral and Brain Sciences*, *40*, e105. https://doi.org/10.1017/S0140525X16000947

Nieh, Edward H., Matthews, Gillian A., Allsop, Stephen A., Presbrey, Kara N.,

Leppla, Christopher A., Wichmann, R., Neve, R., Wildes, Craig P., & Tye, Kay M. (2015). Decoding Neural Circuits that Control Compulsive Sucrose Seeking. *Cell*, *160*(3), 528–541. https://doi.org/10.1016/j.cell.2015.01.003

Nieh, Edward H., Vander Weele, Caitlin M., Matthews, Gillian A., Presbrey, Kara N., Wichmann, R., Leppla, Christopher A., Izadmehr, Ehsan M., & Tye, Kay M. (2016). Inhibitory Input from the Lateral Hypothalamus to the Ventral Tegmental Area Disinhibits Dopamine Neurons and Promotes Behavioral Activation. *Neuron*, *90*(6), 1286–1298. https://doi.org/10.1016/j.neuron.2016.04.035

Nonomura, S., Nishizawa, K., Sakai, Y., Kawaguchi, Y., Kato, S., Uchigashima, M., Watanabe, M., Yamanaka, K., Enomoto, K., Chiken, S., Sano, H., Soma, S., Yoshida, J., Samejima, K., Ogawa, M., Kobayashi, K., Nambu, A., Isomura, Y., & Kimura, M. (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

Noori, H. R., Cosa Linan, A., & Spanagel, R. (2016). Largely overlapping neuronal substrates of reactivity to drug, gambling, food and sexual cues: A comprehensive meta-analysis. *European Neuropsychopharmacology*, *26*(9), 1419–1430. https://doi.org/10.1016/j.euroneuro.2016.06.013

Noritake, A., & Nakamura, K. (2019). Encoding prediction signals during appetitive and aversive Pavlovian conditioning in the primate lateral hypothalamus. *Journal of Neurophysiology*, *121*(2), 396–417. https://doi.org/10.1152/jn.00247.2018

Payzan-LeNestour, E., Dunne, S., Bossaerts, P., & O'Doherty, John P. (2013). The Neural Representation of Unexpected Uncertainty during Value-Based Decision Making. *Neuron*, *79*(1), 191–201. https://doi.org/10.1016/j.neuron.2013.04.037

Pearce, J. M., & Hall, G. (1980). A model for Pavlovian learning: Variations in

the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review*, *87*(6), 532–552. https://doi.org/10.1037/0033-295X.87.6.532

Pessiglione, M., Seymour, B., Flandin, G., Dolan, R. J., & Frith, C. D. (2006). Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature*, *442*(7106), 1042–1045. 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., & Bossaerts, P. (2007). Adding Prediction Risk to the Theory of Reward Learning. *Annals of the New York Academy of Sciences*, *1104*(1), 135–146. https://doi.org/10.1196/annals.1390.005

Preuschoff, Kerstin, Bossaerts, P., & Quartz, S. R. (2006). Neural Differentiation of Expected Reward and Risk in Human Subcortical Structures. *Neuron*, *51*(3), 381–390. https://doi.org/10.1016/j.neuron.2006.06.024

Raji, C. A., Ho, A. J., Parikshak, N. N., Becker, J. T., Lopez, O. L., Kuller, L. H., Hua, X., Leow, A. D., Toga, A. W., & Thompson, P. M. (2009). Brain structure and obesity. *Human Brain Mapping*, NA–NA. https://doi.org/10.1002/hbm.20870

Rangel, A., Camerer, C., & Montague, P. R. (2008). A framework for studying the neurobiology of value-based decision making. *Nature Reviews Neuroscience*, *9*(7), 545–556. https://doi.org/10.1038/nrn2357

Rasmussen, E. B., Lawyer, S. R., & Reilly, W. (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

Richardson, K. A., & Aston-Jones, G. (2012). Lateral Hypothalamic Orexin/Hypocretin Neurons That Project to Ventral Tegmental Area Are Differentially Activated with Morphine Preference. *Journal of Neuroscience*,

32(11), 3809–3817. https://doi.org/10.1523/JNEUROSCI.3917-11.2012

Robinson, Mike J. F., Anselme, P., Fischer, A. M., & Berridge, K. C. (2014). Initial uncertainty in Pavlovian reward prediction persistently elevates incentive salience and extends sign-tracking to normally unattractive cues. *Behavioural Brain Research*, *266*, 119–130. https://doi.org/10.1016/j.bbr.2014.03.004

Robinson, Mike J. F., Anselme, P., Suchomel, K., & Berridge, K. C. (2015). Amphetamine-induced sensitization and reward uncertainty similarly enhance incentive salience for conditioned cues. *Behavioral Neuroscience*, *129*(4), 502–511. https://doi.org/10.1037/bne0000064

Rogers, P. J., & Blundell, J. E. (1984). Meal patterns and food selection during the development of obesity in rats fed a cafeteria diet. *Neuroscience & Biobehavioral Reviews*, *8*(4), 441–453. https://doi.org/10.1016/0149-7634(84)90003-4

Rollins, B. Y., Dearing, K. K., & Epstein, L. H. (2010). Delay discounting moderates the effect of food reinforcement on energy intake among non-obese women. *Appetite*, *55*(3), 420–425. https://doi.org/10.1016/j.appet.2010.07.014

Rushworth, M. F. S., & Behrens, T. E. J. (2008). Choice, uncertainty and value in prefrontal and cingulate cortex. *Nature Neuroscience*, *11*(4), 389–397. https://doi.org/10.1038/nn2066

Samejima, K. (2005). Representation of Action-Specific Reward Values in the Striatum. *Science*, *310*(5752), 1337–1340. https://doi.org/10.1126/science.1115270

Schultz, Wolfram. (2016). Dopamine reward prediction error coding. *Dialogues in Clinical Neuroscience*, *18*(1), 23–32. https://doi.org/27069377

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

Schultz, Wolfram, Preuschoff, K., Camerer, C., Hsu, M., Fiorillo, C. D., Tobler, P. N., & Bossaerts, P. (2008). Explicit neural signals reflecting reward uncertainty. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1511), 3801–3811. https://doi.org/10.1098/rstb.2008.0152

Sharpe, M. J., Marchant, N. J., Whitaker, L. R., Richie, C. T., Zhang, Y. J., Campbell, E. J., Koivula, P. P., Necarsulmer, J. C., Mejias-Aponte, C., Morales, M., Pickel, J., Smith, J. C., Niv, Y., Shaham, Y., Harvey, B. K., & Schoenbaum, G. (2017). Lateral Hypothalamic GABAergic Neurons Encode Reward Predictions that Are Relayed to the Ventral Tegmental Area to Regulate Learning. *Current Biology*, *27*(14), 2089–2100.e5. https://doi.org/10.1016/j.cub.2017.06.024

Shenhav, A., Cohen, J. D., & Botvinick, M. M. (2016). Dorsal anterior cingulate cortex and the value of control. *Nature Neuroscience*, *19*(10), 1286–1291. https://doi.org/10.1038/nn.4384

Siegel, J. M. (2004). Hypocretin (OREXIN): Role in Normal Behavior and Neuropathology. *Annual Review of Psychology*, *55*(1), 125–148. https://doi.org/10.1146/annurev.psych.55.090902.141545

Silva, F. J., Silva, KathleenM., & Pear, J. J. (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, T. V., Boswell, T., & Henderson, L. J. (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, A., & Izquierdo, A. (2019). Adaptive learning under expected and unexpected uncertainty. *Nature Reviews Neuroscience*, *20*(10), 635–644. https://doi.org/10.1038/s41583-019-0180-y

Somerville, L. H., Sasse, S. F., Garrad, M. C., Drysdale, A. T., Abi Akar,

N., Insel, C., & Wilson, R. C. (2017). Charting the expansion of strategic exploratory behavior during adolescence. *Journal of Experimental Psychology: General*, *146*(2), 155–164. https://doi.org/10.1037/xge0000250

Stagner, J. P., & Zentall, T. R. (2010). Suboptimal choice behavior by pigeons. *Psychonomic Bulletin & Review*, *17*(3), 412–416. https://doi.org/10.3758/PBR.17.3.412

Stamatakis, A. M., Van Swieten, M., Basiri, M. L., Blair, G. A., Kantak, P., & Stuber, G. D. (2016). Lateral Hypothalamic Area Glutamatergic Neurons and Their Projections to the Lateral Habenula Regulate Feeding and Reward. *The Journal of Neuroscience*, *36*(2), 302–311. https://doi.org/10.1523/JNEUROSCI.1202-15.2016

Stephens, D. W. (2008). Decision ecology: Foraging and the ecology of animal decision making. *Cognitive, Affective, & Behavioral Neuroscience*, *8*(4), 475–484. https://doi.org/10.3758/CABN.8.4.475

Stice, E., & Burger, K. (2019). Neural vulnerability factors for obesity. *Clinical Psychology Review*, *68*, 38–53. https://doi.org/10.1016/j.cpr.2018.12.002

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: an introduction* (Second edition). The MIT Press.

Sweeney, P., & Yang, Y. (2016). An Inhibitory Septum to Lateral Hypothalamus Circuit That Suppresses Feeding. *The Journal of Neuroscience*, *36*(44), 11185–11195. https://doi.org/10.1523/JNEUROSCI.2042-16.2016

Tajima, S., Drugowitsch, J., & Pouget, A. (2016). Optimal policy for value-based decision-making. *Nature Communications*, *7*(1), 12400. https://doi.org/10.1038/ncomms12400

Takahashi, Y. K., Batchelor, H. M., Liu, B., Khanna, A., Morales, M., & Schoenbaum, G. (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, S. C., Samejima, K., Okada, G., Ueda, K., Okamoto, Y., Yamawaki, S., & Doya, K. (2006). Brain mechanism of reward prediction under predictable and unpredictable environmental dynamics. *Neural Networks*, *19*(8), 1233–1241. https://doi.org/10.1016/j.neunet.2006.05.039

Terrill, S. J., Hyde, K. M., Kay, K. E., Greene, H. E., Maske, C. B., Knierim, A. E., Davis, J. F., & Williams, D. L. (2016). Ventral tegmental area orexin 1 receptors promote palatable food intake and oppose postingestive negative feedback. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, *311*(3), R592–R599. https://doi.org/10.1152/ajpregu.00097.2016

Tordoff, Michael G. (2002). Obesity by choice: the powerful influence of nutrient availability on nutrient intake. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, *282*(5), R1536–R1539. https://doi.org/10.1152/ajpregu.00739.2001

Tordoff, M. G. (2002). Influence of Test Duration on the Sensitivity of the Two-bottle Choice Test. *Chemical Senses*, *27*(9), 759–768. https://doi.org/10.1093/chemse/27.9.759

Tordoff, M. G., & Bachmanov, A. A. (2003). Influence of the Number of Alcohol and Water Bottles on Murine Alcohol Intake. *Alcoholism: Clinical and Experimental Research*, *27*(4), 600–606. https://doi.org/10.1111/j.1530-0277.2003.tb04396.x

Trinko, R., Sears, R. M., Guarnieri, D. J., & DiLeone, R. J. (2007). Neural mechanisms underlying obesity and drug addiction. *Physiology & Behavior*, *91*(5), 499–505. https://doi.org/10.1016/j.physbeh.2007.01.001

van den Berg, P., & Wenseleers, T. (2018). Uncertainty about social interactions leads to the evolution of social heuristics. *Nature Communications*, *9*(1), 2151. https://doi.org/10.1038/s41467-018-04493-1

Wajnberg, E., Fauvergue, X., & Pons, O. (2000). Patch leaving decision rules

and the Marginal Value Theorem: an experimental analysis and a simulation model. *Behavioral Ecology*, *11*(6), 577–586. https://doi.org/10.1093/beheco/11.6.577

Wake, S. J., & Izuma, K. (2017). A common neural code for social and monetary rewards in the human striatum. *Social Cognitive and Affective Neuroscience*, *12*(10), 1558–1564. https://doi.org/10.1093/scan/nsx092

Wang, H., Wen, B., Cheng, J., & Li, H. (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, M., Eshel, N., & Uchida, N. (2017). Neural Circuitry of Reward Prediction Error. *Annual Review of Neuroscience*, *40*(1), 373–394. https://doi.org/10.1146/annurev-neuro-072116-031109

Werlen, E., Shin, S.-L., Gastambide, F., Francois, J., Tricklebank, M. D., Marston, H. M., Huxter, J. R., Gilmour, G., & Walton, M. E. (2020). Amphetamine disrupts haemodynamic correlates of prediction errors in nucleus accumbens and orbitofrontal cortex. *Neuropsychopharmacology*, *45*(5), 793–803. https://doi.org/10.1038/s41386-019-0564-8

Wilson, R. C., Geana, A., White, J. M., Ludvig, E. A., & Cohen, J. D. (2014). Humans use directed and random exploration to solve the explore–exploit dilemma. *Journal of Experimental Psychology: General*, *143*(6), 2074–2081. https://doi.org/10.1037/a0038199

Withrow, D., & Alter, D. A. (2011). The economic burden of obesity worldwide: a systematic review of the direct costs of obesity: The direct healthcare costs of obesity. *Obesity Reviews*, *12*(2), 131–141. https://doi.org/10.1111/j.1467-789X.2009.00712.x

Wosniack, M. E., Santos, M. C., Raposo, E. P., Viswanathan, G. M., & da Luz, M. G. E. (2017). The evolutionary origins of Lévy walk foraging. *PLOS*

Computational Biology, *13*(10), e1005774. https://doi.org/10.1371/journal.pcbi.1005774

Wu, X., Wang, T., Liu, C., Wu, T., Jiang, J., Zhou, D., & Zhou, J. (2017). Functions of Learning Rate in Adaptive Reward Learning. *Frontiers in Human Neuroscience*, *11*, 592. https://doi.org/10.3389/fnhum.2017.00592

Yamanaka, A., Beuckmann, C. T., Willie, J. T., Hara, J., Tsujino, N., Mieda, M., Tominaga, M., Yagami, K., Sugiyama, F., Goto, K., Yanagisawa, M., & Sakurai, T. (2003). Hypothalamic Orexin Neurons Regulate Arousal According to Energy Balance in Mice. *Neuron*, *38*(5), 701–713. https://doi.org/10.1016/S0896-6273(03)00331-3

Yokum, S., & Stice, E. (2019). Weight gain is associated with changes in neural response to palatable food tastes varying in sugar and fat and palatable food images: a repeated-measures fMRI study. *The American Journal of Clinical Nutrition*, *110*(6), 1275–1286. https://doi.org/10.1093/ajcn/nqz204

Yu, A. J., & Dayan, P. (2005). Uncertainty, Neuromodulation, and Attention. *Neuron*, *46*(4), 681–692. https://doi.org/10.1016/j.neuron.2005.04.026