Obesity and environmental uncertainty

Luis Nicolás Luarte Rodríguez

# Abstract

Obesity is a problem of great concern at a global scale, and while its considered a multifactorial disease, the proximal causes are, in great part, due to altered decision making. 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 by framing decision making as developed in its natural environment, and how this can result in unwanted results when obesogenic environments come into play. Additionally, the neural basis of how the proximal and distal can come into play are discussed. The proposed framework could help to understand how the interplay between feeding-related behavior, and the interaction with the environment could lead into obesity.

# Introduction

Obesity prevalence increase in recent in year in addition to its economic cost (**NL7KNKU2#Withrow\_Alter\_2011?**), present a major concern ranging from the society level down to the individual health. Despite of its importance, and the association between impaired decision making and obesity (**UTKKIKVA#Fitzpatrick\_Gilbert\_Serpell\_2013?**; **67T2WUWD#Davis\_Etal\_2004?**), there is not a great wealth of literature including more refined models considering, how decisions about food consumption, could lead to obesity (**A9R3LNZJ#Nettle\_Andrews\_Bateson\_2017?**). Thus, there exists a gap between the proximal causes of obesity, and the decision making processes that lead to it. Moreover, not having a model or theory of the underlying decisions-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 were they most likely evolved, and the environmental setting that shaped them (**P4NSHDMV#Stephens\_2008?**). Foraging theory (**ESYGCSLH#Charnov\_1976?**) offers such possibility as it addresses how animals solve the problem of obtaining sufficient food to survive in environments that are not completely known. As environments characteristics, such as food placement or predator are rarely known, foraging becomes a problem of searching in uncertain environments. While most animals are capable of integrating sensory information and take decision based on that, uncertainty and the computational complexity of such integration, bias organism to more simpler algorithms to pick the optimal option (**9XCDNBAM#Bartumeus\_Etal\_2016?**), because exploitation of environment structural characteristics to generate decision rules that have minimal computational requirements (**WS9KYIYG#Fawcett\_Etal\_2014?**).

The problem setting can be stated succinctly, an animal must satisfy its energetic demands, while searching for food in an uncertain environment, balancing between eating too much (becoming less mobile), and eating less but increasing risk of death upon sudden food scarcity. To solve this animals can exploit that uncertainty of food access predicts food scarcity, and have mechanisms that allow the computation of uncertainty, such as the reward prediction error (**6ZEIWIGL#Colombo\_2014?**).

Understanding how, the previously presented problem, is solved and implemented, offers a framework to explain how obesity is generated under a certain type of environment. To develop such framework, first, how food-seeking behavior relates to uncertainty, and how animals deal with energetic balance is presented. This 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 previous models 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, the properties of obesogenic environments are reviewed in light of the decision models presented.

# Food-seeking behavior and uncertainty

Uncertainty, as broader concept, implies having limited or partial knowledge about something. Here, uncertainty is going to be understood as the knowledge of how a given action, inside a particular environment, relates to outcomes, for example, how does a lever press relates to the delivery of food, pressing the lever always gives food?, or only sometimes. Therefore, uncertainty represents the doubt about getting a food reward, when the lever is pressed, such doubt can 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 (**UKGJNWU7#Preuschoff\_Bossaerts\_Quartz\_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 (**AF2655I2#DeGroot\_Thurik\_2018?**). A particular feature of risk is the inverted ‘U’ shape of its distribution. Where at outcome probability 0 or 1 is at minima, and at maxima when probability is 0.5, this derives from risk being measured as outcome variance between expected reward and actual reward (**UKGJNWU7#Preuschoff\_Bossaerts\_Quartz\_2006?**). If an animal is situated in a natural environment, it is unlikely to have complete information about the action-outcome pairing, and thus is forced to generate estimates of such pairing, 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 (**ESYGCSLH#Charnov\_1976?**). Multiple theories on how an agent must decide to optimally allocate time to each resource patch (**IJSUH426#Wajnberg\_Fauvergue\_Pons\_2000?**), the optimal path to take (**A5N6NMAB#Hills\_Kalff\_Wiener\_2013?**; **M5RXPXSZ#Humphries\_Sims\_2014?**) and evolutionary roots of such strategies (**V2XKZN75#Wosniack\_Etal\_2017?**).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 (**KPZRMRFZ#Rangel\_Camerer\_Montague\_2008?**).

[TAG: foraging in uncertain environments; locomotor activity] If an animal does not have perfect information about its environment, predicting if there is going to be food available in the near future turns into a problem. To avoid such problem, animals can consider how the current food availability fluctuates over time. If fluctuations are big, it indicates high uncertainty over food availability, if small, food availability is certain. How animals could make use of such information? In natural environments, contrary to experimental setups, uncertainty in food availability is probably a direct product of food scarcity. The probability of having a food encounter is reduced proportional to resource levels, up to the point that the levels are so low, that the certainty about resource depletion. Thus, is to be expected that uncertainty effects are in line with food scarcity signaling, that is, a expected reduction in energy expenditure in order to preserve the energetic balance. However, empirical evidence is not so clear in this regard (**U2XEEC7Q#Polo\_2002?**), as both increases and reductions in body mass given increased levels of uncertainty in food availability has been observed (**CFC7DN65#Fokidis\_Etal\_2012?**), and this, might be related to increased locomotor activity (**B3IPWFXL#Ferretti\_Etal\_2019?**). When controlling for total food intake in condition of fixed or variable food availability, birds show decrease their body mass, partly explained by increased locomotor activity (**CFC7DN65#Fokidis\_Etal\_2012?**). Such increase in locomotor activity can be a product of increased foraging bouts (**8RIQXAVT#Lohmus\_Sundstrom\_Moore\_2006?**), or maintaining the amount of foraging bouts, but making them longer (**PN23NXIS#Bordier\_Etal\_2018?**). Additionally, mathematical modelling of foraging behavior, shows that an optimal strategy is to change foraging bouts to the start of the day, probably, in order to account for non-successful foraging bouts (**LLLWQCZE#Bednekoff\_Houston\_1994?**), which has also been observed empirically (**B3IPWFXL#Ferretti\_Etal\_2019?**). These results can be interpreted as stress being a signal of food availability, and the modulation of foraging bouts responding to a decreasing 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 bout (**KA4Y29AG#Kane\_Etal\_2017?**).

[TAG: foraging in uncertain environments; why increase locomotor activity?] 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 as mechanisms to prevent starvation in a near future, and are triggered by food-availability uncertainty, possibly because uncertainty is associated with food scarcity (**C6Z374UG#Anselme\_Güntürkün\_2019?**).

[TAG: foraging in uncertain environments; is it uncertainty or scarcity?] Food scarcity can generate multiple changes, such as the number of foraging spots visited, diet diversity, and others (**G83L8BXA#Harris\_Chapman\_Monfort\_2010?**), which are similar to the ones already discussed. Nevertheless, there exist some differences between food scarcity and uncertainty. A scenario were 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 that this effects are coming from uncertainty rather than scarcity, as changes regarding feeding routine, such as feeders positions, increases food intake compared to the constant environment (**NL4XYLRH#Forkman\_1993?**). (**NL4XYLRH#Forkman\_1993?**) study did not control for food-availability across conditions, however, in both conditions food was enough to satisfy energetic demands. To control for food-availability, equating total time available to get food has been used, and similar effects were reported when comparing predictable and unpredictable settings (**3GQMEPEH#Cuthill\_2000?**).

[TAG: foraging in uncertain environments; resume] Up to this point, there are at least 3 points to consider regarding food uncertainty, (1) under food availability uncertainty, food-seeking behavior is increased (**2YZE6SVW#Robinson\_Etal\_2014?**; **CFC7DN65#Fokidis\_Etal\_2012?**; **U2XEEC7Q#Polo\_2002?**). (2) when uncertainty is increased, strategies to maintain energetic balance, such as hoarding (**C6Z374UG#Anselme\_Güntürkün\_2019?**) or increasing body mass (**BW3RY7GD#Moiron\_Mathot\_Dingemanse\_2018?**; **3GQMEPEH#Cuthill\_2000?**) emerge. (3) such strategies imply a trade-off between preventing starvation and increasing risk of predation (due to reduced mobility), and as such are suggestive of a dynamic balance between increasing body mass to prevent starvation, and not doing so to retain mobility (**8XA5RHVI#Macleod\_Etal\_2005?**).

## Uncertainty increases food-seeking behavior

[TAG: food-seeking behavior in uncertainty; conceptual definitions] 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?. Such behaviors, in experimental settings, are considered as the interactions with cues or apparatuses that are related to reward delivery. When the action is interacting with a conditioned stimuli (such as a lever or light), is called sign-tracking, whereas if interaction is with the food dispenser (or equivalent) is called goal-tracking (**ZW878GVH#Silva\_Silva\_Pear\_1992?**). More specifically, sign-tracking, refers to an approaching behavior towards previously conditioned stimuli and rewards. So, it implies a previous conditional-stimulus and unconditional-stimulus pairing, and, an afterwards tracking of the signal that was previously associated with the reward (**BDFISDRS#Flagel\_2014?**). This distinction is relevant because signal-tracking has been shown to respond robustly to uncertainty in food availability (**S8CHV5KG#Anselme\_Robinson\_Berridge\_2013?**).

[TAG: food-seekign behavior in uncertainty; precisely defining an uncertain environment] When uncertainty is introduced at the stage of conditional and unconditional stimulus pairing, as the probability of reward delivery upon lever pressing, sign-tracking increases as the probabilities of reward delivery approaches 50%, and the amount of reward is more varied (**S8CHV5KG#Anselme\_Robinson\_Berridge\_2013?**). In this case, as the delivery of a given reward gives no information about following one (delivery is determined by a probability function, independent of animal action), it can be assumed that, under Shannon entropy formulation, entropy (which can be understood as a measure of uncertainty) (**X596XYG3#Namdari\_Li\_2019?**) reaches the peak at 50% probability, and, furthermore, it predicts that uniform distributions, with more outcomes, increase uncertainty. Both were the case in the previously presented experiment (assuming uncertainty drove signal-tracking) as 50% probability of delivering 2 or 0 pellets had lower signal-tracking than 50% probability of delivering 0 or 1, 2, or 3 pellets with equal probability (16.7% for 1, 2 or 3 pellets). This, again, points out that increased food-seeking related behavior increases upon increased uncertainty even when food availability is controlled. This effect has been replicated in studies with amphetamine sensitization, where uncertainty (on conditioned stimulus and unconditioned stimulus) and sensitization, independently, augmented sign-tracking behavior, however, the effect of both uncertainty and sensitization was not additive suggesting a ceiling effect (**UZET5L4Z#Robinson\_Etal\_2015?**).

The increased, food-seeking related behavior magnification by uncertainty, has been found with partial reinforcement procedures (**P3MLPUGY#Collins\_Etal\_1983?**), with manipulation of food placement variability (**NL4XYLRH#Forkman\_1993?**), variability on reward quality and delivery delay (**YI6J6Y7I#Craft\_2016?**), and in sequential probability tasks (**YB2MIP4H#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 in a trial by trial basis, because is safe to assume that an agent interested in sensing environment uncertainty does not possess complete information. A plausible mechanism is to sense uncertainty, indirectly, via the reward prediction error (**6ZEIWIGL#Colombo\_2014?**). The reward prediction error is simply

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 (**2BEHEM7X#Sutton\_Barto\_2018?**)

Here the trial series is composed of time steps with a discount factor . The discount factor is there to signal the typical preference for obtaining rewards now rather than latter, on and how big it is will depend on properties of both agent and environment (**R3TZXYBW#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 latter on. However, the main idea is that, over trials, as the expected value approximates the real one, the reward prediction error goes down, nevertheless, if rewards value change, the error goes up reflecting such change (see figure ).

The computation of reward predictions is dependent on the activity of the dopamine system (**8B5TPXB5#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 (**NLDHLRVN#WatabeUchida\_Eshel\_Uchida\_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 (**ZHGB75KH#Bayer\_Glimcher\_2005?**). Around this point two main hypothesis have been formulated, the first, proposes than negative errors are encoded via lowering the fire-rate compared to the baseline (**33K2X73I#Schultz\_Dayan\_Montague\_1997?**). Whereas the second, proposes an opponency system between dopamine and serotonin systems (**JZJAAAYD#Daw\_Kakade\_Dayan\_2002?**). Furthermore, dopamine neurons, in the ventral tegmental area, have been found to encode the future discounted rewards (**JPZD7WDQ#Enomoto\_Etal\_2011?**). This two lines of evidence points that dopamine is capable of encoding expectation, reward value and doing subtraction (perhaps including the serotonin system), showing a significant complexity of this system, which might exceed value-related computations (**698KWSAL#Takahashi\_Etal\_2017?**).

Above, 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) (**AR2TQB84#Fiorillo\_2003?**). Such uncertainty-related signal has also been found in the orbito frontal cortex, amygdala (**YWWID6GW#Schultz\_Etal\_2008?**) and medial frontal lobe (**ZRTDBMRA#Huettel\_2005?**). A plausible hypothesis to link the reward prediction error and uncertainty encoding, can state that, over time, reward prediction error signals are integrated into an uncertainty signal, as, over time, more error is to be expected under higher reward variability (see figure ). However, evidence points towards independent signals of reward prediction error and uncertainty in the orbito frontal cortex (**HVH8KZZ2#Rushworth\_Behrens\_2008?**). Nevertheless, at least at a computational level, the reward prediction error can be used to estimate the reward-related uncertainty (**P2FYNJKR#Soltani\_Izquierdo\_2019?**).

[TAG: prediction error and learning rates, how are these linked?] Prediction error is linked with uncertainty, as the error is expected to increase with higher uncertainty levels. Thus, prediction error allows to asses 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, to reflect this learning rates should be high, so recent prediction errors influence the estimate to a greater measure, whereas if the environment is certain, learning rates should be lower to represent past information (**MYWY9LWC#Wu\_Etal\_2017?**). Considering the development over the course of a task, at the start, higher learning rates are to be expected so learning happens at faster rate, however, as the task is properly learned learning rates should go down, so to not be influenced by random fluctuations (**89DIP32B#EvenDar\_Mansour\_2001?**).

[TAG: prediction error -> uncertainty -> learning rate] 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 (**3B7ZWETH#Preuschoff\_Bossaerts\_2007?**), in the same line, empirical experiments have shown that humans behave according to a reward standard deviation-dependent scaling of reward prediction error (**CMAIA6R7#Diederen\_Schultz\_2015?**), so error should be less impactful when standard deviation is high. This proposed learning rate modifications are in line with the original model by (**NZFTTQJZ#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 (**LIRXSMKQ#Jepma\_Etal\_2016?**), this allows for behavioral flexibility, as large errors signal environmental changes, and increasing the learning rates allow to maximize the importance of newer information. Is important to note that other systems, aside from dopamine, are able to track environment uncertainty level, such as the endocrine system through the stress response, measured as subjetive stress, pupil diameter and skin conductance (**WLQDZ5HW#DeBerker\_Etal\_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 new acquired information must be considered?, a notion to answer this question is that of expected and unexpected uncertainty (**W3DEZPKU#Yu\_Dayan\_2005?**). Expected uncertainty is the variability attributable to the stochastic nature of the reward, whereas unexpected uncertainty assumes that the agent is creating belief about action-rewards associations, and incoming information breaks such beliefs (**9Z525EYW#PayzanLenestour\_Etal\_2013?**). Unexpected uncertainty, thus, has been proposed as top-down process which might be tracked by the Locus Coeruleus norepinephrine activity (**X6X8T8QT#Filipowicz\_Etal\_2020?**; **9Z525EYW#PayzanLenestour\_Etal\_2013?**). Moreover, such activity, measured as pupil diameter, has been found to track the learning rates (**U2GSF8HI#Nassar\_Etal\_2012?**), which represents the end product of assessing 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 (**NZFTTQJZ#Pearce\_Hall\_1980?**), which in turn represent that the current model of the environment must change in order to accurately depict it. Then, the learning rate must increase, so to give more weight to more recent information and correctly update the environment model (**Q3Q977MQ#Faraji\_Preuschoff\_Gerstner\_2017?**).

Up until now, the discussion presented has focused on the calculation of reward value, typically, given a certain action. However, the computation of action-reward value is not directly linked to action choice. Picture a situation where the environment present high levels of uncertainty, and one action, until the present time step, have been associated with high rewards. If we were to choose just based on the maximum reward in such environment, we could miss potential better options, which true value, cannot be appropriately calculated because of variability. Such situation is more specifically defined by the exploration/exploitation dilemma, which posits that an agent, in order to obtain rewards, must ‘exploit’ current knowledge. However, it also must ‘explore’ to determine the best option in the future (**2BEHEM7X#Sutton\_Barto\_2018?**). One of the findings is that the manipulation of dopamine levels modulates striatal representation of reward prediction errors, and subjects treated with L-dopa (a metabolic precursor of dopamine) chose, with more frequency, the option with greater reward compared to the placebo group and haloperidol (dopamine receptor antagonist) group (**87WR6HV3#Pessiglione\_Etal\_2006?**). Although the authors did not report the temperature parameter (the one that determines the balance between exploration and exploitation) of the model, given that in all three group the optimal option was learned, it can be interpreted that the increase in dopamine levels effectively induced a bias toward exploitation. Direct evidence on the effect of L-dopa in exploration/explotation parameters, effectively shows that exploration is reduced, and this is associated with modulation of uncertainty signals in the insula and anterior cingulate cortex (**8FIHXB6G#Chakroun\_Etal\_2020?**).

## Assessing uncertainty in the present and future

Although the factors determining obesity as an outcome are multiple (**UQ9I8YFE#Ang\_Etal\_2013?**), it is reasonable to assume that the more immediate cause is excess intake relative to energetic demands. Moreover, excess intake is determined in an instance to instance basis, where a decision considering short and long-term benefits/risks must be made. With this in consideration, one can assume that obesity, in part, is caused by sub-optimal short/long-term benefit/risk assessments when making feeding decisions. If this was the case, areas that are related to computing options value in the short/long term, such as the ACC, should be in some way impaired.

Delayed discounting refers to the depreciation of a certain reward as a function of the time required to obtain it (**9YRMQU6H#DaMatta\_Gonçalves\_Bizarro\_2012?**). As such, it provides measures of how reward-related systems bias decision to the short or long term. Obese subjects show a robust tendency to steeply discount future rewards (**FCVQ7SB6#Amlung\_Etal\_2016?**), thus, favoring short-term rewards.

Furthermore, ACC, among other structures, shows relative atrophy in obese subjects (**THS94923#Wang\_Etal\_2017?**; **5FBDWF55#Raji\_Etal\_2009?**), suggesting an impairment of the previously mentioned functions. These findings can be interpreted as if impairment in environment uncertainty assessment results in a preference for short-term rewards. If this were the case, palatable food sensory cues, which trigger food-intake, would dominate over more long-term modulated decisions, such as healthy food intake (**J37T9F5T#Higgs\_2016?**).

Higher future rewards discounting paired with increased motivation to work for food, predict higher caloric intake (**GHWUZPNL#Rollins\_Dearing\_Epstein\_2010?**), and this effect seems to hold even for low energy-density food (**TLRSYNN6#Epstein\_Etal\_2014?**). The rate of reward discounting, thus, informs about the predisposition to increased energetic intake, independent of possible food-property related effects. Similar effects have been found in children (**8QS2B4AF#Best\_Etal\_2012?**), but not in adult males (**IATULRKH#Smulders\_Boswell\_Henderson\_2019?**). Moreover, these effects seem to be directly related to body fat (**7B4XUYMB#Rasmussen\_Lawyer\_Reilly\_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.

denotes the estimated value at a given state, and is used to model the agent learning rate. Additional parameter has been proposed to model sensitivity to reward (**JGKMEKV6#Huys\_Etal\_2013?**; **GU4KJGAS#Kroemer\_Small\_2016?**), such that the temporal difference error accounts for the subjective value of obtained rewards.

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 (**GU4KJGAS#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 (**DBGUDZRL#Mathar\_Etal\_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 (**WCFW3TEU#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 (**PE46B3Z8#Kube\_Etal\_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 (**67GSY6VY#Horstmann\_Fenske\_Hankir\_2015?**). Taken together, this finding suggests that obesity overfeeding is not only reliant on increased reward sensitivity (more reward sensitivity is assumed to increase intake), but other parameters such as learning rates can determine the overall valuation of the reward, biasing decision-making to immediate rewards, that paired with highly palatable food can lead to excess caloric intake. This, because, while palatable food definition is not standardized (**NBA5EBE8#Fazzino\_Rohde\_Sullivan\_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 (**LFEELSI4#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 (**2BEHEM7X#Sutton\_Barto\_2018?**). As presented previously, the reward prediction error representation is able to guide the chosen policy of the agent (**87WR6HV3#Pessiglione\_Etal\_2006?**). Heuristics, which are strategies that rely heavily on exploiting environment statistical properties, have been proposed to be guiding decision-making in uncertain environments (**J7XQF8LY#Hafenbrädl\_Etal\_2016?**). Some heuristics are thought to be an evolutionary derivative from uncertain environments (**S53QRP94#VanDenBerg\_Wenseleers\_2018?**). However, what aspect of uncertainty is the one used to selected the optimal policy is not clear (**LFIY67XG#Gershman\_2018?**). (**LFIY67XG#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 more salient signatures represented the best fit. The first model fit corresponded to Upper-Confidence-Bound (**NXA58KQ4#Auer\_CesaBianchi\_Fischer\_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:

Here represent the expected value of taking action , in the denominator represents how many times action has been chosen up to a certain point , as time appears in the numerator, when action is not chosen its upper-bound will increase, but decrease if such is continously chosen. Note that (**LFIY67XG#Gershman\_2018?**) used a modified version of this algorithm to reflect human decision stochasticity, nevertheless, this provides enough insight into how it considers uncertainty. Author found indirect support for this policy, by considering that reaction times are faster when estimated rewards are more different (**2V3MQIX5#Tajima\_Drugowitsch\_Pouget\_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 and according to:

Where is the model expected reward, and 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 is updated as and , when the observed reward is a loss (or ‘0’) and . Findings by (**LFIY67XG#Gershman\_2018?**) noted that signatures corresponding to this model were that choice stochasticity (exploration actions) were proportional with the level of uncertainty in the model distributions.

Evidence for directed exploration, based on uncertainty levels, has been found in humans (**AFIWCGVS#Blanco\_Sloutsky\_2019?**; **VRHNLHDG#Wilson\_Etal\_2014?**). However, while directed exploration seems to be a robust strategy in humans, certain aspects emerge and vary throughout the life span (**26M86HJZ#Somerville\_Etal\_2017?**), pointing towards a complex and dynamic system. The main idea behind the previously presented models are 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 (**BUXRNM89#Mikhael\_Bogacz\_2016?**) (see figure ) and represents it at a neural level (**AR2TQB84#Fiorillo\_2003?**). Finally, action policies are intimately related to uncertainty, thus establishing a clear link between feeding-related behavior and environment uncertainty.

## Uncertainty representations at the neural level

If uncertainty can modulate food-seeking behavior in order to increase intake and better sustain energetic reserves, it is expected to have at least 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 has been found to be involved in decision-making actions such as action selection (**6JBBK7KQ#Balleine\_Delgado\_Hikosaka\_2007?**), action value representation (**5CLQA59J#Samejima\_2005?**), and reward representation (**UVXRU55G#Wake\_Izuma\_2017?**). Markov decision tasks can be used to test behavior under certain or uncertain environments. Such task comprises of a set of states, which are accessible by performing certain 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, whereas the dorso lateral prefrontal cortex showed greater activation under the certain condition (**G5PZHPCI#Tanaka\_Etal\_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, thus, making 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 (**S53QRP94#VanDenBerg\_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 (**V9CIKRBE#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 (**9SRBHI4W#Davis\_Choi\_Benoit\_2010?**), effectively signaling environment dynamics.

As previously stated, environment dynamics need to be taken into account in order to appropriately interpret obtained rewards. If I visit a restaurant and the food served is delicious, my rating of the restaurant should not be too hasty as this could be just good luck. However, if this has always been the case, giving a high rating would be the correct choice. In term of rewards, uncertainty is high when a given rewards give no information about the ones to come, conversely, certainty is achieved when a given reward gives all information about the following one. Direct tracking of environment volatility has been found to be well represented in the ACC (**BHR2NAEI#Behrens\_Etal\_2007?**), presumably by encoding some sort of learning rate that bias valuation of rewards more to the short-term if volatility is high, and to the long-term is volatility is lower. The competing hypothesis of ACC describes its function to a decision-difficulty sensing unit, or demand of control when overriding default action is more optimal (**J9QC5JYH#Shenhav\_Cohen\_Botvinick\_2016?**). However, it should be noted that (**BHR2NAEI#Behrens\_Etal\_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 the uncertainty of a given environment, an agent must pair the value obtained with the action performed. For each action possible, the agent updates the value of the action-reward tuple based on the reward prediction error.

Temporal dynamics of action-reward pairing and reward prediction error are such that the former occurs first relative to the later. Such temporal difference is reasonable because the pairing should be represented when taken action, and the prediction error requires feedback in order to compare obtained versus expected rewards. Considering this, the action-reward pairing has been found to be correlated to activity at the putamen, whereas rewards-prediction error, to be represented in the caudate nucleus (**DAKDZEVY#Haruno\_Kawato\_2006?**). However, as the authors point, both structures are likely to be involved in a larger loop containing the ACC, which would make sense to integrate reward evaluation over states, actions, and environmental uncertainty, and optimally influence following behavior.

It can be inferred from the way action-reward pairing is stated that it corresponds to action selection based on a history of rewards, which are mediated by the reward prediction error. Inhibition of putamen activity has effectively shown a reduction in performance when the task requires the consideration of reward history to select correct actions (**XZBUNRIS#Muranishi\_Etal\_2011?**). Signal encoding, however, seems to be more complex, as basal ganglia direct pathway encode rewards outcomes, and the indirect pathway represents the next-action selection (**4F2PUL7U#Nonomura\_Etal\_2018?**). Together, this points to a multi-structure network that represents expected and obtained rewards as an error, which allows easing computational requirements as the current state needs only to be compared with the expectation, that encompasses all previous history of rewards. Moreover, this signal updates rewards given actions, while considering environment volatility and the proper weighting of immediate versus long-term rewards. Thus, allowing to optimize behavior even when environments are non-stationary and rapidly changing.

### Orexin role in reward function

Orexin is a neuropeptide with two isoforms, orexin-A and orexin-B, mainly associated with regulation of arousal and feeding behaviors (**PQWIJHAP#Siegel\_2004?**). While orexin neurons are mainly localized in the lateral hypothalamus, they project widely throughout the brain (**C4QSBWHV#Nambu\_Etal\_1999?**; **TUU47UNG#Harris\_AstonJones\_2006?**; **PQWIJHAP#Siegel\_2004?**). Lateral hypothalamus and orexin have been related to feeding and reward-related behavior (**S6TNU7XM#Harris\_Wimmer\_AstonJones\_2005?**). Lateral hypothalamus regulates feed behavior (**MW2YHV9D#Margules\_Olds\_1962?**; **QWCCGZTT#Sweeney\_Yang\_2016?**; **LJVQFQ6K#Anand\_Brobeck\_1951?**), and does this by controlling orexin activity, where increased orexin activity increases feeding (**QFWKWDA5#Ardianto\_Etal\_2016?**; **TPGJBDDD#Kotz\_Etal\_2002?**). Furthermore, orexin-related arousal increase is modulated by extracellular glucose concentrations, and neuroendocrine markers of energetic balance(**KPIHYUYF#Yamanaka\_Etal\_2003?**), pointing to the inclusion of ‘contextual’ information to its function. On the other hand, the reward-related learning and valution seem to be in charge of the dopamine system, and significantly sustained by the activity in the ventral tegmental area [(**8TKQDEML#D?**)‘Ardenne\_Etal\_2008; (**R775Y2Q6#Eshel\_Etal\_2016?**)], that uses the reward prediction error to promote task learning (**E6MZ8HIA#Keiflin\_Etal\_2019?**). While, ventral tegmental area, is involved in the encoding of more that prediction errors, this might be more complex as to include current beliefs about the environment (**SZZJ3VQF#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 hypothalus encodes contextual information, (2) can rely 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 (**V2PX8TRL#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, while preserving reward sensitivity (**BGNECK7I#Bayard\_Etal\_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 (**IQWZPMP7#Burdakov\_2019?**). Such function of lateral hypothalamus is well supported by its anatomical disposition (**SC7QH8NN#Geisler\_Zahm\_2005?**).

#### Lateral hypothalamus relies 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 (**S3LMSWG5#Geisler\_Etal\_2007?**), connectivity, with the same type of activity, exist between the lateral hypothalamus and lateral habenula, which may exert indirect control over the ventral tegmental area (**SZMQLQTN#Stamatakis\_Etal\_2016?**). In addition to glutamatergic activity, those projections also contain a GABAergics element, which is activated by leptin action (**GDVXZGJY#Leinninger\_Etal\_2009?**). Among those innervations, a significant portion contains orexin neurons, and is mainly limited to the ventral tegmental area (**4PLXQQT9#Fadel\_Deutch\_2002?**). These projections have an important functional role as they are involved in compulsive sucrose seeking (**HT5LXV4U#Nieh\_Etal\_2015?**), and more general behavioral activation throughout inhibitory GABAergics outputs from the lateral hypothalamus to the ventral tegmental area (**GZ3AZ6UU#Nieh\_Etal\_2016?**).

The lateral hypothalamus orexin activity inputs into the ventral tegemental area via orexin receptor type 1 (**KN9TLADJ#Richardson\_AstonJones\_2012?**), which allow 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 (**XIJMRM6S#Baimel\_Etal\_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 tegemental area is open. Furthermore, it points that such information might modulate structures beyond the ventral tegemental area.

#### Lateral hypothalamus modulates the prediction error in the ventral tegmental area

Orexin activity increases the intake of palatable food when injected in the ventral tegmental area (**TESJ97B4#Terrill\_Etal\_2016?**; **Q8IHJK68#Mattar\_Etal\_2020?**), moreover, the orexin role in the lateral hypothalamus and ventral tegemental also modulated more general reward-related behavior (**S6TNU7XM#Harris\_Wimmer\_AstonJones\_2005?**; **YMQYAEQ8#Mahmoudi\_Etal\_2020?**). Furthermore, activation of the lateral hypothalamus neurons projecting to ventral tegmental area promotes reward seeking behavior, and displays a prediction error-like profile activity (**HT5LXV4U#Nieh\_Etal\_2015?**). Lateral hypothalamus representing a prediction-error like activity, and its connectivity to the ventral tegemental area, prompts to think of a direct modulation of the dopamine prediction error.

Lateral hypothalamus stores, previously learned, stimulus-reward pairings (**4ESME7SG#Sharpe\_Etal\_2017?**; **V2PX8TRL#Noritake\_Nakamura\_2019?**), this can then inform about the expected value of a given action or stimulus in an previously experienced environment. If such information is not available the prediction error magnitude should increase as proper expectation is never properly formed. An interesting case is the one where stored expected values are withdrawn at the time of prediction error computation, but kept stored for the expected value update. (**4ESME7SG#Sharpe\_Etal\_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 lead to increased performance, whoever faster convergence can lead to suboptimal decision making in more complex scenarios (**89DIP32B#EvenDar\_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 (**9MIJEAWP#AstonJones\_Etal\_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 canonical role in feeding (**2ZT7CAFB#Jennings\_Etal\_2013?**; **V8LPCV6E#Delgado\_Anand\_1952?**) 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 (**7SLC53RQ#Werlen\_Etal\_2020?**).

For orexin to modulate the reward function, specifically the prediction error, a series of conditions were considered, as the capabilities of the lateral hypothalamus to encode contextual information, and effectively input that into the ventral tegemental area. Moreover, it was considered that such input should be capable of altering the prediction error signal computed in the ventral tegmental area. All such capabilities are sustained by anatomical and functional evidence, and as such posit the role of orexin in a pivotal role to the computation and update of rewards expected value. The main importance of assessing the previously mentioned requisites, is to add an important part to the presented models, that is, 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 importance to surivival. 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) (**VTI2FVJG#Brunstrom\_Cheon\_2018?**). Sustaining an insurance against scarcity strategy (increasing caloric intake) for prolonged time periods, assuming sufficient caloric availability and equal energetic expenditure, leads to obesity, which is specifically defined by the percentage of body fat (**BFYKZGMM#Bhadoria\_Etal\_2015?**). If such insurance strategy were to only accurately respond to environmental uncertainty, being introduced to a high caloric density environment, should cease such strategy, however, this often no the case. Moreover, caferia-style diets, where standard diet is supplemented with food rich in carbohydrates and/or fats, robustly generates obese animals (**EBJDV6NP#Leigh\_Kendig\_Morris\_2019?**; **C6PMZNGR#Hariri\_Thibault\_2010?**). 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 (**AAD8FGG7#Trinko\_Etal\_2007?**; **4JUZW55S#GarcíaGarcía\_Etal\_2014?**; **VWAZILDA#Filbey\_Myers\_Dewitt\_2012?**; **HM4RM78V#Noori\_CosaLinan\_Spanagel\_2016?**; **GDQBDJJ2#Carlier\_Etal\_2015?**). 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 as usually observed with psychostimulants exposure, and increases amphetamine self-administration compared with groups exposed to certain food delivery (**ELZ2PYYM#Mascia\_Etal\_2019?**). In natural situations, uncertainty, usually, translates to food insecurity, which refers to having food-availability affected by physical, economical or other related factors (**X8KX96JF#Barrett\_2010?**). Food insecurity has been related to increased BMI in wealthier countries, with higher risk of obesity in women (**4LJKTR3N#Moradi\_Etal\_2019?**), the increased association in women has been explained in terms of higher risk when having lower energetic reserves (**A9R3LNZJ#Nettle\_Andrews\_Bateson\_2017?**). However, how the association is present in wealthier countries results of more interest to consider obesogenic environments. A simple explanation considers that wealthier countries have more readily access to food, so uncertainty-derived motivation to eat leads to effectively eating more (**A9R3LNZJ#Nettle\_Andrews\_Bateson\_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. Now, how this contextual information comes into play when taking decisions, that will eventually lead into obesity?.

Cafeteria-type diet and high-fat diets induce obesity (**CAGXZWY4#Rogers\_Blundell\_1984?**; **9I3RBWWB#Buyukdere\_Gulec\_Akyol\_2019?**), an intuitive interpretation is that both diets contain energetically dense food, so every feeding-bout results in a increased caloric intake, which over-time would lead to obesity. However, a complementary analysis on 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 food, whereas high-fat diets usually consists on fewer types of food, considering this the diversity in cafeteria diets might induce obesity more effectively than fewer food item high-fat diet (**9I3RBWWB#Buyukdere\_Gulec\_Akyol\_2019?**). This effects has been observed across multiple scenarios including more variety on the food or substance of interest (**D6LH64XI#Dao\_Etal\_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 (**C6Z374UG#Anselme\_Güntürkün\_2019?**), and that a given value is affected by other value representations in the following way:

This essentially show that a given option value is normalized by (1) the amount of other options, and (2) by the value of such options. represents the firing rate value encoding, while is the currently encoded option, and are the other options. Parameters , , and are used to generate realistic firing rates (**S8GZWPNK#Louie\_Khaw\_Glimcher\_2013?**). In model simulations, and monkey models of choice behavior (**S8GZWPNK#Louie\_Khaw\_Glimcher\_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 amount of options makes the values more similar. Previous model could be a potential mechanism to explain why increasing food variety or food options 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. Is important to note that mechanisms underlying intake increase due to variety are not know, and more ecologically-leaned explanations are available (**YD49TU6B#Tordoff\_2002?**).

# Conclusion

Uncertainty can be estimated by the ventral tegemental 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 raises 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 posible reason of the increased intake when food is more varied, or when many options are available. Obesogenic environments, aside from being dense in energy rich foods, makes food value representations fuzzier, and, due to noise in value encoding, 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, alternatively, (2) noisier reward value estimation lead to either increased exploration to obtain a better estimate, or increased prediction errors, due to worse value estimation increasing the learning rate as if uncertainty was high, biasing decisions toward short-term rewards. Ultimately this consist of hypothesis of the proximal mechanisms resposible for increased intake in obesogenic environments, and ecological (**YD49TU6B#Tordoff\_2002?**), and foraging theory (**ESYGCSLH#Charnov\_1976?**) could provide explanations based on distal mechanisms, such as sustainable resource exploitation.

# Figures

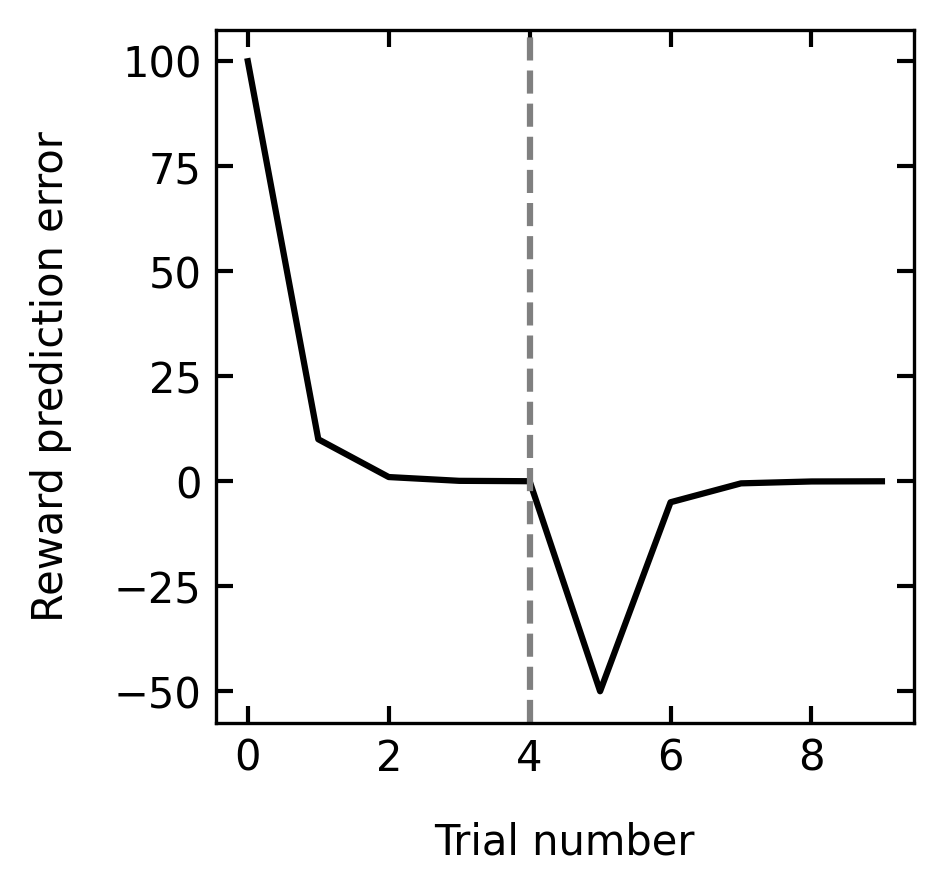
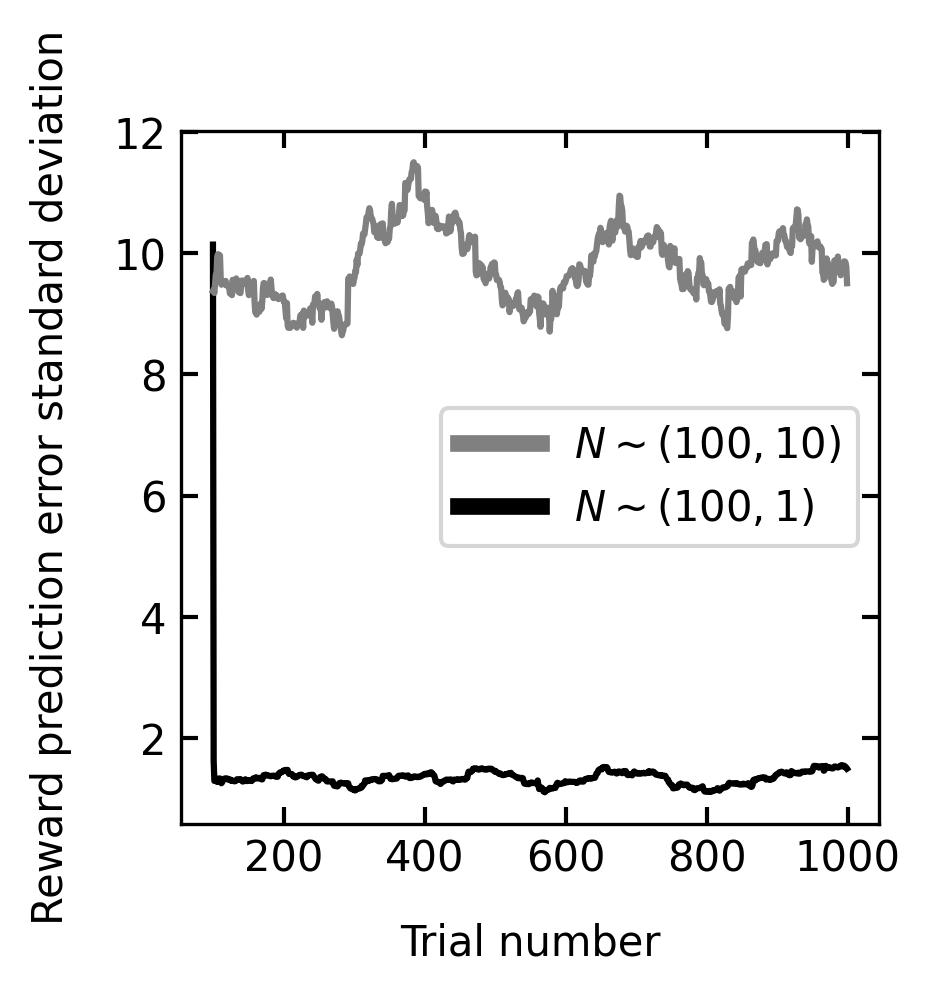
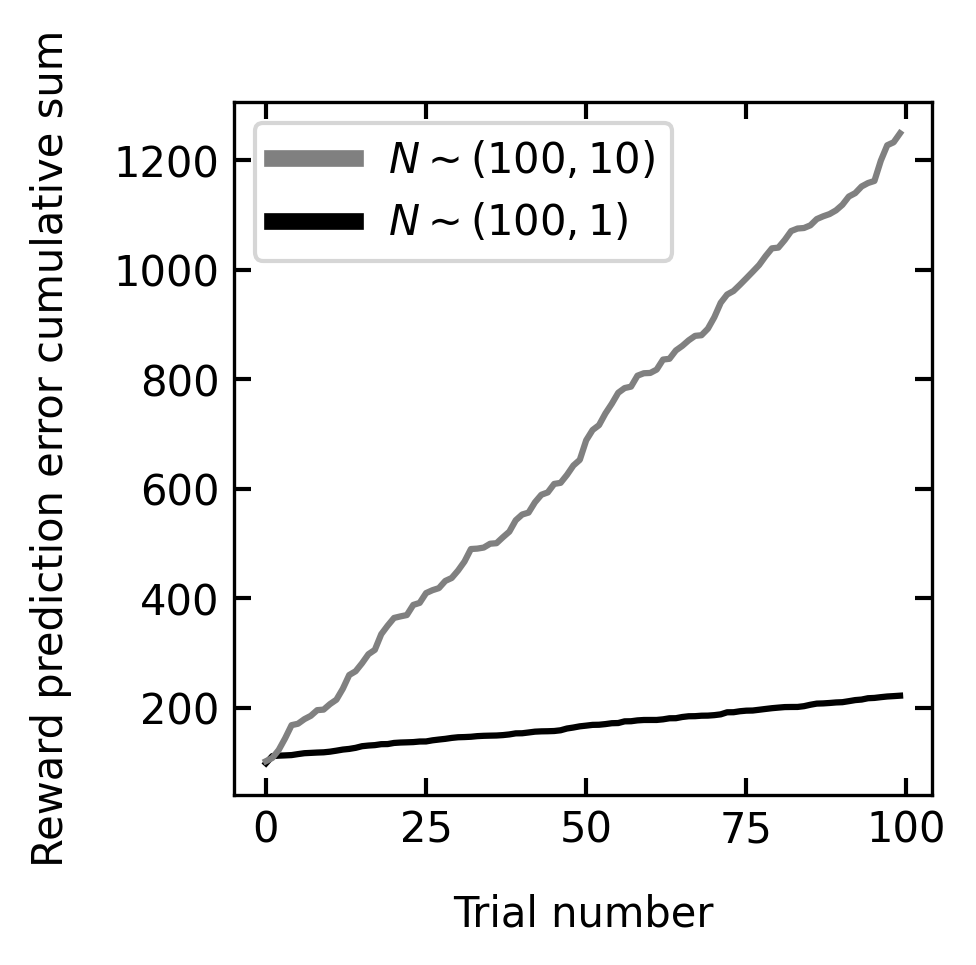


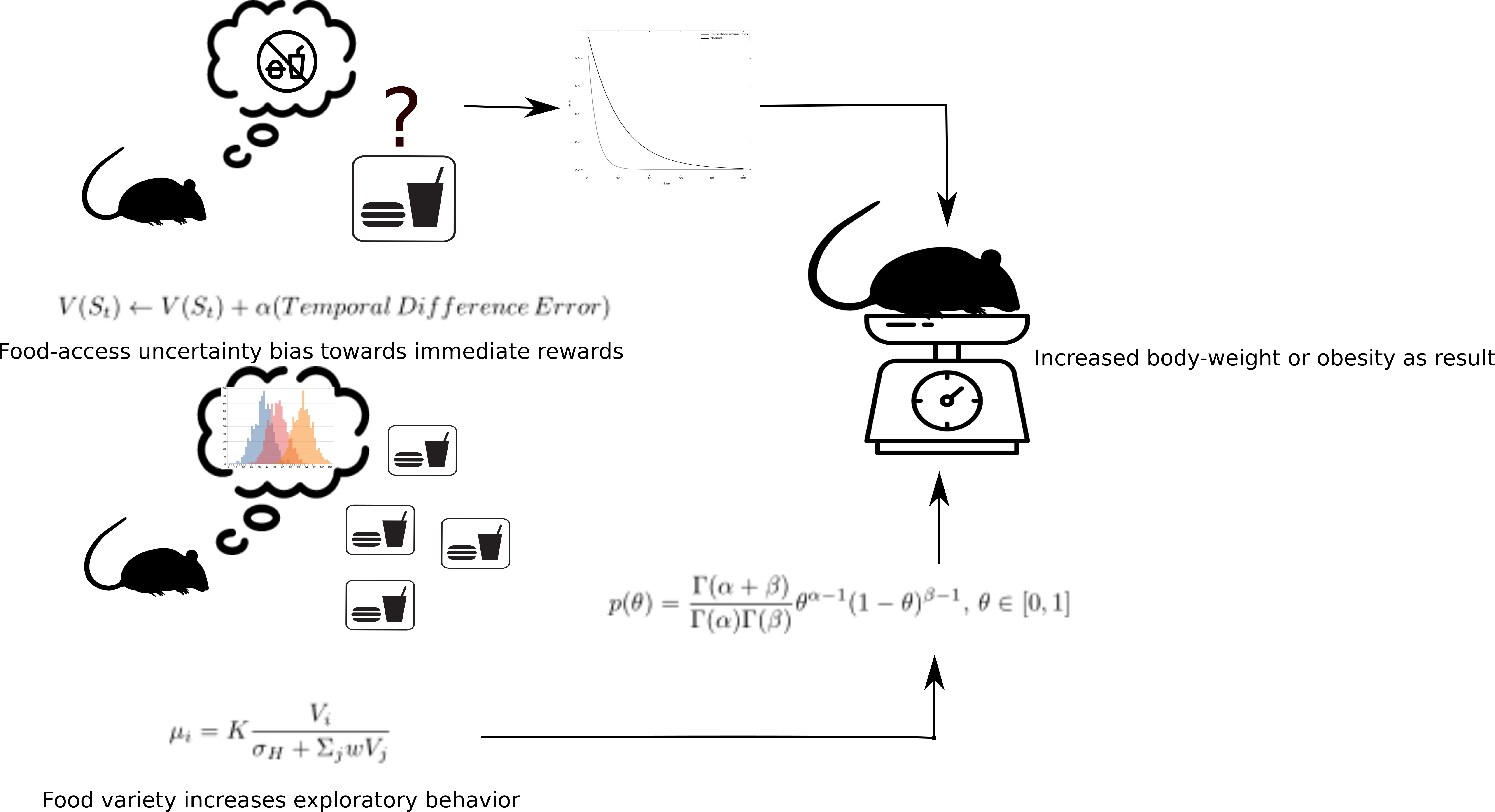
Figure shows a series of 10 trials, where from trial 0-4 the true reward value is 100, and for the remaining trial its 50. A very basic agent was simulated to update its estimates based on the reward prediction error. Initial estimates were set at 0. Notice how during trial 0-4 reward prediction errors 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



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 to the underlying uncertainty of the distribution (using the standard deviation as measure of uncertainty).



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 proportional 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 reward true value, however, in the more uncertain environment this will happen at a slower pace.



Graphical abstract

# References