Jennifer Nagel, University of Toronto, March 29, 2023
Forthcoming in: *Putting Knowledge to Work: New Directions for Knowledge-First Epistemology*
Edited by Arturs Logins and Jacques-Henri Vollet, Oxford University Press.

Natural curiosity

1. The intrinsic desire to know

Knowledge is a naturally attractive state. As Aristotle saw it, "all men by nature desire to know" (*Metaphysics* A1, 980a21), but men are not alone in this. Curiosity, still characterized by at least some contemporary researchers as "the intrinsic desire to know" (Gottlieb & Oudeyer, 2018, 764), is evident in humans of all sorts from early in infancy (Liquin & Lombrozo, 2020); it is also said to appear in a broad array of other animals, including monkeys (Wang & Hayden, 2019), various types of birds (Auersperg, 2015), rats (Small, 1899), and octopuses (Byrne, Kuba, & Griebel, 2002). Even fruit flies show behavior that is taken to suggest a rudimentary form of curiosity (Lewis, Negelspach, Kaladchibachi, Cowen, & Fernandez, 2017).

To say that curiosity is an intrinsic desire for knowledge is to say that curious creatures in some sense want knowledge for its own sake, and not just for its contributions to securing other primary rewards, such as nourishment. However, one might wonder whether curiosity, so understood, could really be present in such a broad range of creatures. Why exactly would an octopus need an intrinsic desire for knowledge, if that is indeed the best way of explaining its exploratory and playful behavior? Even if there are many things that octopuses must know in order to thrive, one might wonder why they couldn't learn these things through processes of trial and error driven by simpler incentives, such as hunger. Indeed, the suggestion that non-human animals are pursuing knowledge for its own sake may sound somewhat ridiculous.

Nevertheless, several lines of evidence arguably point in that direction. Two important strands are outlined in a classic paper of Daniel Berlyne's, a paper that begins with the cautionary epigram, "Animals spend much of their time seeking stimuli whose significance raises problems for psychology" (Berlyne, 1966: 25). First, Berlyne observes that playful and exploratory behavior emerges too early in life to be explicable by an association with the satisfaction of other basic needs. Second, he notes that animals will sometimes attend to novel and complex stimuli even when it is hazardous to do so, and even in preference to satisfying immediate thirst or hunger. In his view, curiosity seems to have some serious independent weight of its own, a finding that he describes as "rather embarrassing" to early twentieth-century theories of motivation, theories which aimed to explain behavior strictly in terms of the pursuit of physical gratification and the avoidance of physical harm (Berlyne, 1966: 25).

Despite very substantial progress, some embarrassment lingers around curiosity today, with psychologists noting its importance and pervasiveness while lamenting a lack of scholarly consensus concerning its fundamental nature and purpose (Dubey & Griffiths, 2020; Kidd & Hayden, 2015). One source of difficulty concerns curiosity's target (or targets). Peter Carruthers, who accepts that many animals are curious, has argued that a desire for knowledge should be impossible for animals lacking the metacognitive ability to conceptualize that mental state (Carruthers, 2018). Carruthers now sees animal curiosity as guided by a kind of 'model-free metacognition', but he wavers (in ways to be examined in section 3) on whether the target of this guidance is knowledge or mere belief (Carruthers, 2023: 15, 16).

Berlyne himself would not have described nonhuman animals as seeking knowledge. He draws a line between perceptual curiosity, described as aiming at "information-bearing stimulation" (Berlyne,

1966: 31), and epistemic curiosity, described as aiming at knowledge. In his theory, nonhuman animals have perceptual but not epistemic curiosity; humans have both. However, epistemologists may wonder whether this result can be traced back to Berlyne's decision to define knowledge as "information stored in the form of ideational structures" (1966: 31); perhaps a better way of understanding knowledge would have animal perceptual curiosity turn out to be knowledge-oriented as well. Meanwhile, among contemporary empirical researchers, many take all curiosity to be aimed at knowledge (Gottlieb & Oudeyer, 2018; Kang et al., 2009; Kobayashi, Ravaioli, Baranès, Woodford, & Gottlieb, 2019; Loewenstein, 1994), while others speak in terms of drives or desires for information (Hsee & Ruan, 2016; Oudeyer & Smith, 2016). Still others combine the two, for example by defining curiosity as "a reward-learning process of knowledge acquisition or information seeking" (Lau, Ozono, Kuratomi, Komiya, & Murayama, 2020: 531), or by characterizing curiosity as "a drive state for information" whose purpose is "to motivate the acquisition of knowledge and learning" (Kidd & Hayden, 2015: 450, 457). The relationship between knowledge and information is typically left unexplained in this literature, with the rare efforts to spell it out – "We define knowledge as information that is internal to the agent" (Silver, Singh, Precup, & Sutton, 2021: 5) – raising at least as many philosophical questions as they answer.

If the idea of knowledge as an internal agential state sounds unorthodox, there are other moves in this literature which might seem to lead us even further outside familiar epistemological territory. One influential paper on intrinsic motivations such as curiosity characterizes them as "based on mechanisms that drive learning of skills and knowledge, and the exploitation and energisation of behaviours that facilitate this, on the basis of the levels and the variations of such skills and knowledge directly detected within the brain" (Baldassarre, 2011: 3). The suggestion that levels of knowledge can be directly detected within the brain is surprising, and may sound especially jarring to 'knowledge-first' epistemologists who see knowledge as a factive mental state, a state whose objective, reality-matching correctness is its essential core (Williamson, 2000). This sort of passage might create the impression that empirical research on curiosity uses the word "knowledge" to pick out something quite different from whatever is meant by that word in epistemology, and that it will be hard to connect epistemological and empirical work in this area.

This chapter builds a fresh defense of the theory that curiosity is an intrinsic desire for knowledge, exactly by connecting knowledge-first epistemology to empirical work on natural curiosity. If it seems at first that empirical researchers and epistemologists are doomed to talking past each other, a closer look shows that their theories can be brought into contact, to the benefit of both sides. Knowledge-first epistemology can deliver a sharper theoretical characterization of the target of natural curiosity, and show how it is possible for animals to be curious without being reflective about their mental states. At the same time, empirical models of natural curiosity can clarify what it even means for knowledge to be a factive mental state, a state of a type that can only bind an agent to truths. Research into animal curiosity, together with related models in reinforcement learning, shows how knowledge can function as a special kind of adaptation of the agent to reality, an adaptation resulting in a type of mental state whose very existence depends essentially on the truth of its contents. Curiosity accelerates this adaptation by making it a direct goal for the agent, rather than a mere byproduct of the search for other goals. However, it will take some work to explain how this happens, and just why our adaptation to reality needs acceleration.

It will also take some work to clarify the relationship between information and knowledge. As an initial step, it may help to focus on an example of Jürgen Schmidhuber's (2010), involving a vision-based agent who encounters a television screen broadcasting random white noise. The display is

terrifically information-rich in the classic signal-detection-theoretical sense (Shannon, 1948): at each moment, the precise array of pixels on screen is one of a vast range of possible arrays, and entirely unpredictable from everything that has preceded it. Regular patterns have redundancies that allow a compressed representation to transmit them with fidelity; the maximal disorder in the white noise means that it cannot be compressed, and can only be transmitted or reproduced exactly by means of a file with the same large size as the original. If curiosity were just a drive to encounter raw information, this visual buzz should be a source of endless fascination for a curious agent. The reason that it is not, Schmidhuber suggests, is that there are no patterns that the agent could make progress in compressing, no regularities that any agent could come to know. He applies this idea equally to what happens at the other end of the information spectrum, where a constantly darkened room affords minimal information. The dark room is also boring, but now because its regularity is learned at once, affording no further progress. What curious agents seek is not just raw information, but information at a level that poses an appropriate challenge to their current cognitive powers. Characterizing that level will turn out to be a surprisingly difficult problem, and a problem central to understanding the nature of curiosity.

The white noise example does not show that cognitive scientists are wrong to describe curiosity as involving information-seeking; rather, it raises questions about what sort of information is sought, and how. To answer these questions, we will need to look at a larger framework incorporating the relationship between the powers of the agent and the features of the environment producing informative signals, plus the structures in the brain producing metacognitive signals about the agent's level of adaptation to reality. With this larger framework in view, I will argue that the best way of making sense of curiosity is as a state aimed at knowledge gain. A quick advance sketch of the main ideas may be useful here. First, because our most basic form of learning is a kind of prediction-error correction, creatures like us gain knowledge when we are surprised, when events violate our expectations. (Note that the televised static which does not enable us to learn anything is information-rich but unsurprising, as we have no particular expectations about its configuration from one moment to the next.) Following a cue from research in Reinforcement Learning, we will examine the proximal signals behind natural curiosity, and formulate it as a kind of appetite for surprise. Surprise functions as a reward for the curious creature (in a technical sense of "reward" that will be explained), but because creatures like us gain knowledge from surprising situations, curiosity amounts to a desire for knowledge gain, or so I will argue. The resulting theory has broad applicability, because even animals lacking a reflective understanding of knowledge can still feel surprise. Curiosity is not strictly necessary for knowledge gain: a creature who must interact with the environment to satisfy basic extrinsic needs such as hunger and thirst will also learn at times when events violate its expectations. However, creatures who are curious benefit from an interaction between their reactive prediction-error correction processes and the active surprise-seeking force of their curiosity; this internally adversarial interaction accelerates knowledge gain in ways that are very helpful for biological agents in environments like ours.

After making a case that the best way of making sense of empirical research on curiosity is to understand this state as aimed at knowledge gain, I will need to explain what is going on talk of detecting knowledge levels in the brain. Here I will situate curiosity among other motivational forces which also make use of proximal mechanisms in the brain to secure distal aims; when speaking loosely, cognitive scientists may sometimes fail to distinguish between markers of knowledge and knowledge itself, but this is not to say that the distinction is actually dispensable. But before tackling questions about the mechanisms that enable curiosity, and more abstract questions about the structure of curiosity itself, we can get a better sense of the target phenomenon by reviewing some

data on how curiosity shows up in the natural world. The next section surveys research on apparently curious behavior in various types of animals, including humans.

2. Apparently curious behavior in animals

The curiosity of the most-studied animal in comparative psychology has been remarked upon from the discipline's beginning. In his 1899 study of the development of the white rat, Willard Small insists that curiosity is "the most striking" of the animal's intellectual traits, emphasizing that "it is really curiosity, which is customarily spoken of as boldness by writers upon the natural history of the rat," and insisting that this curiosity develops early: "by the time they are three weeks old it is inordinate and overbalances fear" (Small, 1899: 99). In his view, "rats manifest curiosity apart from that curiosity which is directly associated with nutrition and reproduction" (Small, 1899: 90). This last observation has held up well under subsequent research. Exploratory behavior in the rat is not simply driven by immediate hunger, because rats will actively explore a maze even if they have just been fed. Indeed, the urge to explore is strong enough that hungry rats who are returned to their cages after being deprived of food for 24 hours will ignore openly available food if the bedding has been refreshed in their absence, and first "give themselves up for a while to explorations over and through their new bedding" (Dashiell, 1925: 208). Rats will also cross a floor that delivers a painful electric shock in order to explore new areas, and will do so when they are neither hungry nor thirsty (Warden, 1931).

Exploratory behavior is not just random activity. Rats who have never been fed outside their cages will explore mazes systematically, showing a strong preference to visit the options where they haven't most recently been: if they are relocated back to an earlier juncture, they significantly favor the right branch of a Y-or T-maze if they had previously explored the left, and vice versa, as opposed to choosing at chance (Dennis 1934). This tendency for 'spontaneous alternation' seems to be driven by the relative novelty of the place, for the rat, and not simply by the novelty of the motor activity of turning in a different direction. When these factors are disentangled, for example in mazes where the right and left fork swiftly lead to the same destination, rats no longer strongly prefer to avoid the path just taken (Montgomery, 1952; Sutherland, 1957). Exploration driven by relative novelty is better than random movement for foragers in a changing world, not only because it produces more efficient spatial discovery, but also because the longer one has been absent from a place, the greater the chance some food has appeared in one's absence.

Rats can be taught specific routes, or sequences of right and left turns, to gain extrinsic reward, but curiosity prompts the development of a more powerful form of spatial navigation, a cognitive map. Where meaningful routes have designated start and goal points, and a value contingent on the reward of reaching that goal, maps can be pressed into service from any direction, with no goal fixed in advance, enabling navigation along paths never taken before; in addition, maps are more robust in the face of environmental perturbations, where specific routes collapse at the introduction of an obstacle (O'Keefe & Nadel, 1978).

The cognitive maps developed by rats are maps in a broad sense of the word: they encode not only the spatial layout of reality, but also its causal features (Tolman & Brunswik, 1935; Wikenheiser & Schoenbaum, 2016). To flesh out the causal side of the map, rats seem to exhibit curiosity in their motor interactions with material objects, exploring new materials unprompted and without any apparent prospect of material reward. The benefits of this exploratory behavior become evident when novel problems must be solved. An illustrative example can be found in a study of the cognitive differences between rats raised in plain cages, with nothing beyond bedding, food and

water, and rats raised in cages enriched with wooden toys, sections of tunnel, and metal enclosures (Renner, 1988). Researchers placed both types of rat in an arena with an escape route concealed under an obstruction box that could be entered only from the top; to simulate predation, the rats were chased by a noisy remote-controlled car. On their first time in the arena, all rats explored, but only the rats who had prior experiences of exploring materially enriched environments were able to escape through a cardboard ramp in the obstruction box within the three-minute experimental time limit (Renner, 1988: 52). Just as prior spatial exploration leads to mapping of a type that enables future navigation, prior material exploration equips rats with greater courage and skill in future interactions with various materials. These representations formed through these two types of exploration guide responses to unforeseen challenges and opportunities. Structurally similar types of cognitive mapping apply across a great range of task domains, including abstract non-spatial domains, and appear in a great range of animals, including humans (Behrens et al., 2018; Whittington, McCaffary, Bakermans, & Behrens, 2022).

Rats may show especially conspicuous curiosity, but many animals will approach and manipulate novel items. An early study of over 200 species of captive zoo animals investigated their responses to a range of objects: wooden blocks, steel chains, wooden dowels, rubber tubing and crumpled paper, scaled roughly to the animal's size, and introduced to the animals' cages for a series of six-minute test sessions (Glickman & Sroges, 1966). These stimuli provoked active responses across taxonomic groups, with primates most engaged, and carnivores close behind. The reptiles were less engaged, the authors note, "with the notable exception of an Orinoco crocodile who lunged at, pushed, and bit all of the test stimuli" (Glickman & Sroges, 1966: 164). It is hard to extract clear lessons about animal curiosity from this work, however: one might worry, in particular, that these particular human-selected items were shaped or structured to lend themselves better to the manipulation and chasing behavior of primates and carnivores, while other objects might have been more enticing to the lethargic snakes and armadillos. (Anatomical fit did make sporadic differences across animals: for example, anteaters scored high amongst the "primitive mammals" because they could insert their long tongues into the rubber tubing.) Even within a given species, there was considerable variation: one hedgehog scored zero in responsiveness, only bristling at the introduction of the objects, while another hedgehog chewed vigorously on the blocks and tubes and carried them around the cage. The same prehensile-tailed porcupine who was unresponsive to all the objects in the first test period grasped and chewed all of them in the second, while dancing on his hind legs (Glickman & Sroges, 1966: 173-4). This diversity in responses raises many questions: for example, one might wonder about the extent to which heightened responsiveness could have been a sign of confusion over whether the provided objects were edible, or dangerous, or perhaps some misapprehension interacting with the animal's current state of hunger, arousal, or fatigue.

More systematic and controlled investigation can resolve some of these ambiguities. Octopuses swiftly approach every object dropped into their surroundings, a tendency noticed by Aristotle, who saw it as a weakness: "The octopus is a stupid creature, for it will approach a man's hand if it be lowered in the water" (*History of Animals* VIII(IX).37, 622a3-4). Contemporary theorists now draw quite a different conclusion, taking this behavior as evidence of the octopus's cognitive energy. Research on captive octopuses confirms that they do uniformly begin by approaching any novel object; one study tried edible shellfish versus similarly-sized cube- and snowflake-shaped plastic toys, and found the octopuses approaching everything dropped into the tank within minutes (Kuba, Byrne, Meisel, & Mather, 2006a). Following the initial approach, most of the time, objects were explored with a tentacle or two, and sometimes fondled into the central web near the mouth. Food was more likely to win tactile exploration when the octopus was hungry as opposed to sated, in

contrast to the toys, which were explored equally often, and with frequent touches, whether or not the octopuses were hungry. Equal attention to the toys in the hungry and sated conditions is one sign that the tactile exploration of the toys is not simply 'misplaced predation' by an animal confused about what it can eat. Signs of curiosity continued after the initial tactile exploration. After initial touching, food was either eaten or ignored, but the toys attracted ongoing attention. Most of the octopuses in a related study carried on to such play behaviors as passing the plastic toys repeatedly from arm to arm, and towing them across the water surface (Kuba, Byrne, Meisel, & Mather, 2006b). It is natural to wonder what octopuses are gaining from this behavior, why exactly they are motivated to invest energy into their interactions with inedible and practically useless pieces of plastic, and indeed why they risk approaching everything in the first place, even (potentially octopus-hunting) human hands. Doubtless some benefits offset the costs. If their exploratory behaviors work to expand their general practical competence, it is noteworthy that this competence ends up being considerable: octopuses have demonstrated the capacity to solve strategic problems calling for significant behavioral flexibility, for example where an extended series of steps must be performed correctly to gain food reward (Richter, Hochner, & Kuba, 2016).

Primates also show strong tendencies towards sensorimotor exploration. Given an elaborate mechanical puzzle, solvable only by manipulating a series of six catches, clasps and levers in the right order, rhesus monkeys manipulate it avidly until they can solve it with high accuracy, producing a learning curve that "does not appear to differ in any way from a typical learning curve obtained on animals under hunger or thirst motivation" (Harlow, 1950). This sequencing of fine motor skills is learned with no reward other than the presumed satisfaction of finding a reliable way to open the innermost metal flap, in a device that could be seen from the outset to contain nothing edible.

On the side of theoretical knowledge, curiosity in primates and other animals can be shown by their willingness to pay for information with no clear strategic value. Gambling paradigms have been used to show that rhesus monkeys are willing to forfeit some material reward (juice) to gain useless information about chance outcomes (Wang & Hayden, 2019). Strikingly, the amount of juice forfeited was proportional to the amount of information gained, with willingness to pay more where the chance outcome was closer to a 50-50 gamble for the monkey, the peak of informativeness in the formal entropic sense (Shannon, 1948).

Humans show some similar tendencies, in ways that are easier to probe because humans can explicitly rate their level of curiosity. Humans report higher curiosity to see the early resolution of gambles with greater outcome uncertainty (van Lieshout, Vandenbroucke, Müller, Cools, & de Lange, 2018). Indeed, we are willing to endure physical pain to see trivial but highly chancy matters resolved. In one ingenious experiment (Hsee & Ruan, 2016), participants were left in a room with a box containing some trick pens that would deliver a mildly painful electric shock when clicked, amid normal pens. Participants who were ostensibly waiting for the start of another study were told that they could click the pens to kill time, if they wished. Ten of the pens were marked with a red sticker, ten with a green sticker, and ten with a yellow sticker; the research assistant mentioned that the red stickers indicated a live battery which would always produce a painful shock on clicking the pen, the green stickers indicated a dead battery with no risk of shock, and the yellow ones were mixed. Left to their own devices for four minutes, participants spontaneously clicked more of the yellow-sticker uncertain-outcome pens (M=4.16, SD=3.67) than both of the other colours put together (green: M=1.69, SD=2.29; red: M=1.03, SD=1.79). The unpredictability of the yellow-sticker pens seems to be part of their allure: those who simply wanted to alleviate boredom through variety could more easily have done so by systematically selecting green or red pens at will (Hsee & Ruan, 2016: 664).

The study is an elegant demonstration of the force of pure curiosity, both because the participants are willing to endure pain, while apparently gaining nothing of practical value by clicking the yellow pens, and because the green and red pens are tangible control stimuli whose value seems to differ only in its predictability. It is not easy to explain what is going on here while keeping in mind that curious humans do not feel a chronic temptation to flip coins.

Recent work has revealed deep similarities in the neural processing of human cravings for useless knowledge and for food. To elicit curiosity, participants can be asked a trivia question or shown a brief video of a magic trick, and then offered a chance to learn the answer or find out how the trick is done. The brain regions activated by such offers match those activated when hungry participants are shown an image of a snack and offered a chance to win it (Lau et al., 2020). In both cases, participants are then willing to take some risk of an electric shock to be told the answer or given the snack, accepting greater risk where the reported curiosity or craving is stronger. Decisions about these gambles for food or for trivial knowledge play out in the same striatal reward areas of the brain (Lau et al., 2020).

One robust pattern in human curiosity deserves special mention. Curiosity generally seems to be higher for questions on which we have some middling level of confidence (Loewenstein, 1994). This "inverted-U-shape" curve shows up in multiple domains. For example, trivia questions produce lower reported curiosity among those who report either very low or high confidence of getting the answer right; peak reported curiosity appears in most people for questions on which they report a level of uncertainty near the midpoint (Kang et al., 2009). Something similar appears in the attentional patterns of infants: they are most attracted to moderately challenging visual events, tending to look away from anything too simple or too complex, where complexity is a function of the predictability of the event, given the infant's prior experience (Kidd, Piantadosi, & Aslin, 2012). This "Goldilocks effect" of seeking partial familiarity is seen as a way of keeping cognitive resources engaged in the zone between what is already known and what is currently unknowable, the zone most promising for progress in learning.

3. Carruthers and the metacognitive challenge

Having surveyed apparently curious behavior in a wide array of animals, the next task is to defend a unified account of this behavior. The view that curiosity is an intrinsic desire for knowledge has met with criticism from a philosopher who also sees it as widespread among animals. Peter Carruthers is ready to credit even bees with curiosity, showing up in their looping exploratory flights, for example. He has argued against "metacognitive" views of curiosity on the grounds that they demand too much self-reflection on the part of the curious animal. To desire something, he proposes, a creature must have at least some concept of it: so for example, "in order to want food, one must be capable of identifying (some) foodstuffs" (Carruthers, 2023, p.5). On Carruthers's fairly minimal notion of what is needed for concept possession, bees can have concepts like FOOD and HOME, but they lack the kind of model of their own minds needed for a concept like KNOWLEDGE or BELIEF. Bees can be curious without having the power to represent states of knowledge, he contends, because they have motivational structures that tend to produce knowledge while bypassing the need to model it as a target. Following Dennis Whitcomb (2010) and Jane Friedman (2013), Carruthers sees curiosity as a motivational state whose contents are questions, rather than propositions: the question picks out a set of alternatives, and the question-directed motivational state triggers behavior that settles the question in favor of one of them. We can describe these structured sets of alternatives as whatever is systematically designated by question words such as WHERE, WHICH, WHEN, and WHAT; Carruthers proposes that these question concepts fall within the cognitive capacity of very simple creatures. We

could think of the WHERE concept, for example, as structuring the spatial composition of the bee's cognitive map of the environment. So, in Carruthers's view, the lost bee is not flying in search of something it represents as a change to its own mental state; rather, its exploratory flight is motivated by a question-embedding attitude, with a first-order content like *where the hive is*; likewise, "a monkey looking to see which of two hiding places contains food is motivated by a state with the question *where the food is,* or perhaps *which of these contains food*" (Carruthers, 2023: 5).

Carruthers's first (2018) version of this view involved a stark denial of metacognition in animal curiosity. As he saw it then, curiosity simply "recruits and motivates actions that have been sculpted by evolution and subsequent learning to issue in new knowledge (moving closer to the target of the question, looking at it, sniffing it, and so on). And just as fear can be apt to issue in safety without representing it (rather, safety is the normal effect of running away), so curiosity can be apt to issue in new knowledge without representing it" (Carruthers, 2018, p.9). In this edition of the anti-metacognitive view, the satisfaction of curiosity (coming to know an answer to the question) does not figure in any way in the bee's motivational profile; curiosity is simply a motivational force that ends up having learning as a result.

Carruthers has since moved a step closer to those who take curiosity to involve a representation of knowledge: he now holds that animal curiosity is produced and sustained by signals with the content *not known* and *learning is happening* (Carruthers, 2023). The modification is driven in part by the observation that subjective experiences of satisfaction are crucial to motivated learning: if hunger enables animals to learn new ways of getting food, this is because the state of hunger is aversive, and eating is especially pleasurable for a hungry animal. New behaviors that satisfy hunger are reinforced by these hedonic rewards. Likewise, "affective conditioning of successful curiosity-satisfying behavioral strategies can only happen when reward signals are created by learning the answer" (Carruthers, 2023: 15). Of course, the monkey who wonders which of two hiding places now contains food will gain (food) reward from investigative behavior that (also) results in knowledge, but this way of improving the animal's investigative behavior is already available without curiosity. The search for food is a straightforward instrumentally-motivated inquiry; by contrast, curiosity impels the hungry rat to turn down the hedonic promise of openly available food until it has thoroughly explored its fresh bedding. But the sheer fact that such non-instrumental exploration will answer a question (WHAT is there? –shredded paper) cannot release the needed hedonic signal unless the animal has some way of monitoring the satisfaction of curiosity.

The challenge for Carruthers is to explain the availability of signals for *not known* and *learning is happening* without positing some representation of knowledge in the animal. His inventive solution is what he calls 'model-free metacognition',[1] in which an ignorant animal's ambivalent reaction to an unknown stimulus itself comes to signal that the stimulus is unknown, and the resolution of this ambivalence is reward. There is precedent for this approach in other work on animal metacognition: Nate Kornell, in particular, has argued that animals can make decisions about their levels of confidence in gambling tasks on the basis of cues such as their own wavering or response times (Kornell, 2014). Animals do not need to engage in any introspection in order to respond systematically to such cues; their decision behavior in gambling tasks can be shaped by features of

[1] The next section will explore what it means for learning to be model-free (as opposed to model-based) in more detail; the main point of contrast for present purposes is that model-free cognition is habitual, requiring no representation of the goal state, where model-based cognition can also be strategic.

their own cognitive performance in ways that correlate nicely with their accuracy, without the animals needing to consult some inner model of this correlation.

To illustrate his theory of curiosity, Carruthers gives an example of a cat confronted with a new mechanical toy: the cat's attention is drawn by the noise of the toy, and a prediction error signal results when the cat fails to recognize it. The toy's size and motion will partially activate rival neural populations associated with roughly similar familiar dynamic objects (MOUSE, BALL, etc.) alongside a neural population representing NOT KNOWN. Each of these competing neural populations activates certain motor responses: "The MOUSE-representation will activate the motor processes involved in stalking, pouncing, and biting; the BALL-representation will activate sequences of patting-and-chasing; and the PAPER-ON-STRING-representation will activate chasing and biting. Yet in competition with all these will be the investigative actions distinctive of curiosity (attending, approaching cautiously, patting). Since the representations underlying the latter build strength to the extent that the others don't (and because the item is not recognized), it carries the information *not known*. And given the role of that information in stabilizing the behaviour in question, and rendering it adaptive, that is what it represents, too"(Carruthers, 2023: 24-5). This *not known* signal has both indicative and directive content, Carruthers observes.

Carruthers sees curiosity as always (or almost always) initiated by prediction error signals (2023: 21), and sustained by the *not known* signal until either the animal's question is answered as a result of the investigative behavior (the bee spots home and switches to its straight-line homing flight) or some other priority emerges. But epistemologists might wonder what it means to answer the question. At times, Carruthers frames things in terms of knowledge: "Curiosity is only satisfied in a way that is adaptive (in the way that it has been designed by evolution to be satisfied) when the agent comes to know some fact that answers the question embedded in the state of curiosity" (2023: 15). But he also uses disjunctive formulations involving belief and learning: "Curiosity about what something is will include within its content that one acquires a belief of the form *that is an X,* or that one learns what the thing in question is" (2023: 16). If curiosity is essentially a response to the shuffling ambivalence generated by an unrecognized object, then the cessation of that ambivalence could indeed be produced either by learning (coming to know) what the object is, or by simply forming a belief (even a false belief) about its identity. We could try to reconcile these two ways of talking by reading the first (knowledge-based) formulation as about the idealized natural function of curiosity, and the second as about its ordinary manifestation in us, where beliefs falling short of knowledge can indeed satisfy us at times, but not in the manner that nature intended. However, Carruthers elsewhere doubles down on the belief formulation: "States of belief figure among the referential satisfaction conditions of curiosity. It is belief-acquisition of the right sort (matching or answering the question) that removes curiosity and serves as curiosity's reward, as well as being curiosity's adaptive function." (2023: 16). Advocates of the view that curiosity aims at knowledge could grudgingly agree that states of belief figure among its satisfaction conditions, just on the strength of the entailment between knowledge and belief. But if all it takes for a belief to be "of the right sort" is that it matches or answers the question, then this characterization of the adaptive function of curiosity really is at odds with the earlier characterization in terms of knowledge. False answers and coincidentally correct answers are still answers, matching the form of the question. If we identify mere belief-formation as the target of curiosity, it becomes somewhat harder to see its adaptive value: when animals enter into states of ambivalence, curiosity drives them to rehearse the kinds of actions that liberate them from these states and make them decisive again. But if their decisiveness or uncertainty reduction does not necessarily tend to steer them towards the truth, curiosity loses some of its appeal: in general, boldness might not be a benefit.

Of course, there are a number of ways out of this problem: for example, Carruthers could observe that natural selection will have ensured already that sensorimotor engagement with things will generally improve the accuracy of our judgments about them, so any struggle with a question is likely to leave us better off, epistemically. But an account of curiosity that does not make deliberate use of the special features of knowledge (as opposed to belief) may be missing something important, and Carruthers's account of curiosity does seem to leave a number of questions unanswered.

The way Carruthers describes it, curiosity is a reaction to perturbations of our usual homeostasis into ambivalence, which we are then keen to resolve by whatever innate or learned behaviors have resolved such ambivalence in the past. This does not quite seem to capture the way in which curious animals actively search out trouble, crossing electrified floors or plowing through the hidden reaches of their bedding, notwithstanding the openly visible bowl of food that one might expect to cue a more decisive and unproblematic motor response. It is not obvious that anything unexpected in the bedding needs to trip off a prediction error to motivate the curious search. More generally, we can become acutely curious about things that are outwardly familiar: the sealed envelope that one has promised not to open still looks exactly the same, sitting on the hall table three days after its arrival, while curiosity about its contents continues to build. One might also wonder about the extent to which curiosity triggers new behaviors, as opposed to reviving innate or previously learned ones; here we might think about the rats' material exploration, the anteaters' insertion of their tongues into those rubber tubes, or the octopuses' experiments with dragging plastic toys across the water surface.

To the extent that Carruthers sees curiosity as activating prior learned or innate patterns of investigation, he faces a special challenge in explaining how exactly curiosity-driven inquiries differ from instrumental inquiries. We can grant that knowledge is important for animals while still wondering why exactly they need a special motivational state whose adaptive function is to expand knowledge, as opposed to learning what they need to know on the basis of simpler motivations such as hunger. The latter type of learning could be enough to support even quite complex intelligent behaviors through secondary reinforcement.

According to Carruthers, there is no sharp line here: "the core difference between intrinsically motivated states of questioning like curiosity and interest, on the one hand, and instrumentally motivated inquiries, on the other, is that the former are appraised against long-standing interests and values, whereas the latter are appraised against current goals"(2023: 26). In an example he uses to illustrate this point, a visitor to a park might have her attention drawn by animate motion in a tree. Supposing that she is broadly interested in wildlife ("she likes to watch nature programmes and cares about the diversity of the natural world") and initially fails to recognize the creature, this will spark an affective state of curiosity in which she draws closer to examine it. Without that interest, a person might disregard the disturbance and continue walking. As Carruthers sees it, "Curiosity and interest are sustained by signals with the content *not known*, then. But such signals on their own are insufficient. One's ignorance also needs to be appraised as somehow relevant to one's goals or underlying values and interests." (2023: 25).

The example is a compelling one, but one may wonder whether it is missing some of the peculiarity of curiosity. Is the research participant who clicks the yellow-stickered pen something of a pen-fancier? What goals or underlying values drive the monkey who spends hours manipulating Harlow's empty metal device, or the octopuses with their plastic toys? These activities seem strangely independent of the ordinary goals of the agents involved; perhaps this independence from

instrumental reward is an important feature of what makes curiosity its own intrinsic form of motivation, and not just a diluted or more diffuse version of ordinary learning. Carruthers does have a point that not just any experience of something as unknown will trigger curiosity, but here we might want an account with a greater capacity to predict just what will do the trick. One way to explain his wildlife case would be to return to the observation that curiosity is highest where we have partial familiarity, or more generally, that it is triggered by stimuli that present a manageable learning challenge. There is, after all, something that the yellow pen-clicker can easily learn, so perhaps what drives him is the abstract shape of the learning situation, rather than a mild interest in pens or shocks. A broad interest in a topic could suffice to produce partial familiarity, but perhaps it is the availability of a certain sort of challenge that really matters here: perhaps curiosity drives us to learn in a way that is responsive to the level of our cognitive resources.

4. Curiosity in reinforcement learning

Research in the branch of artificial intelligence known as Reinforcement Learning (RL) has clarified the importance of marshalling cognitive resources well. In RL, artificial agents[2] discover through trial and error how to act in ways that tend to yield reward in artificially designed environments; as these environments become more realistic and complex, it becomes increasingly important for the agents to make efficient use of their capacities to absorb and process information. In research on humans and other animals, it can be hard to tell which aspects of intelligent behavior are innately specified, and which are learned from experience; by contrast, RL research provides a cleaner slate for measuring differences between ways of controlling action, and for testing the impacts of supplying the agent or the environment with various features.

One feature that can be added to the agent is some kind of intrinsic motivation for gaining knowledge. There are a number of ways of doing this (for a partial taxonomy, see Oudeyer & Kaplan, 2007); here I will not attempt an exhaustive survey, but instead highlight the approaches that are most relevant to a better understanding of natural curiosity. Our initial puzzle will then take a clearer form: if even fruit flies show signs of curiosity, we might wonder whether it is somehow essential to agency, or essential to gaining knowledge of an environment. I will argue that it is not; completely incurious RL agents can gain knowledge incidentally, in pursuit of other objectives, and in some environments these incurious agents can perform very well, even optimally. If curiosity-free knowledge acquisition is possible for agents motivated only by the pursuit of non-epistemic reward, we can wonder exactly what agents ever gain from the addition of intrinsic motivation to pursue knowledge directly. To answer this question, we will look at environments and tasks where curious RL agents radically outperform their incurious counterparts, and we will see some commonalities with the natural world, and with the sorts of tasks that biological agents must succeed at, in order to thrive.

RL research does not shy away from talk of knowledge; for example, an important class of RL agents is characterized as having "an adaptive world model … reflecting what's currently known [by the agent] about how the world works," together with "a learning algorithm that continually improves the model (detecting novel, initially surprising spatio-temporal patterns that subsequently become known patterns)." (Schmidhuber, 2010: 1). Granted, there are a few researchers who deny that RL

[2] Readers who think that only biological agents can have mental states can still see artificial agents as useful models of phenomena such as mental state possession; ultimately, humans and other animals also learn through similar processes of reinforcement. For simplicity, this chapter follows the common practice in RL of attributing mental states to artificial agents, rather than speaking of them as modelling mental state possession.

agents can literally learn or come to know anything. For example, Selmer Bringsjord and colleagues, who start from the premise that knowledge is true belief for which the subject can produce an explicit justification, contend that RL agents do not gain knowledge so conceived, and conclude that such agents do not learn in the literal sense of the word (Bringsjord, Govindarajulu, Banerjee, & Hummel, 2017). However, one way of seeing what is wrong with this way of proceeding is to consider its application to the human case: as Cameron Buckner puts it, "This benchmark produces the surprising verdict that children do not really learn how to walk, talk, or recognize objects" (Buckner, forthcoming-a: 11). If we want to keep the idea that learning is literally knowledge gain, an epistemology that fits more smoothly with RL is the theory according to which knowledge is a mental state that is essentially factive, the type of state whose correctness is necessary for its existence. To see that RL agents can successfully achieve this kind of cognitive adaptation to their environments,[3] we can start by taking a closer look at how they are supposed to learn (this quick summary follows the 2020 edition of Sutton & Barto's textbook *Reinforcement Learning*; for a more detailed tour of the philosophically interesting features of RL, see Haas, 2022).

In RL, causal influence runs both ways between the agent and the environment: the environment influences the agent by supplying a series of observations, and the agent influences the environment by producing a series of actions. The environment could be artificially simple (say, a tic-tac-toe game whose states are grid configurations of x's and o's) or arbitrarily complex. Observations are a signal of the state of the environment, sometimes a very direct and complete signal, such as board positions in a game like chess,[4] sometimes something more complex, such as the arrays of pixel values in consecutive frames of a video game. Given these observations, the agent computes the current state at every turn. This is a trivial computation in the tic-tac-toe environment, where the state is identical to the current observation, and more difficult in richer environments, where patterns of past observations are needed to differentiate superficially similar situations which are fundamentally distinct (like similar-looking doors leading to different rooms) and to compress superficially different situations which are fundamentally similar (like different-looking doors leading to the same room). Because of inevitable limitations in memory, time and computational power, agents in complex environments cannot maintain a separate representation for each state, but must instead discover classes of relevantly similar states through processes of generalization, discounting noise in the observational signal from the environment, and discarding features of observations that are irrelevant to action.

An action can be a move that will simply determine the next state that the agent will experience (such as taking one step forward in a simple maze); but more generally, state-action pairs set a probability distribution over the possible next state outcomes (like pressing the trigger on a weapon that

[3] RL agents could know *more* than what they learn from experience: they could reasonably be seen as having some innate knowledge supplied by their designers, such as knowledge of the rules of a game like chess. Some of the initial (genetic) programming of biological agents could also constitute (innate) knowledge of certain stable features of the environment, where the accuracy of this programmed state is essential to its existence in the agent through natural selection. However, questions about innate epistemic states are set aside for present purposes; in what follows I focus on learning from environmental interaction within the lifespan of the agent.

[4] Because chess has rules about the repetition of board positions, and about the one-time availability of castling actions, the state of the environment is not fully expressed in a single observation of the current board position, but only in the set of current and prior positions. That richer set of observations does however carry complete information about the state of the environment. Many environments are unlike chess in being only partially observable; however, even partially observable environments like the natural world have regularities that are knowable for an agent with adequate powers of memory and computation.

sometimes misfires). Actions are selected, in each computed state, according to a rule known as the policy of the agent. The agent's policy dictates how it will act in any computed state by setting some probability distribution over the available actions for that state (so even a random selection of actions in every state could count as a policy, and indeed this might be a fine initial policy for RL agents starting to learn a new task). As time elapses, the agent's actions lead to new situations supplying fresh observations, perhaps including an experience of reward, and new possibilities for action.

The experience of reward plays a decisive role in refining the agent's policy over time. Reward is a special type of observation, a scalar signal that serves as a target for the agent, like points in a game for a player who has the sole objective of maximizing his score.[5] Patterns of behavior that yield positive reward are reinforced. The reward signal can be distributed densely or sparsely among the other observations: for example, in video games where the current score is displayed as a running tally, the score might either change frequently as time elapses, or it might sit at zero until very substantial progress has been made. Either way, the reward signal dictates what the agent will learn to do, although learning will generally be quicker in the presence of a dense reward signal, where steady "breadcrumbs" of reward frequently update the agent's evaluations of possible actions. (Curiosity will turn out to be especially important in scenarios with sparse extrinsic reward.)

A central puzzle in goal-directed action is how future reward can have a bearing on the present moment of choice. In RL this puzzle is solved by the agent's evolving representation of value, a measure of the extent to which the immediately available actions in each state have historically tended to produce reward. The value of an action in a given state is not the reward this action immediately produces, but the total average long-term reward it can be expected to lead to, conditional on the agent's policy, taking past experience as a guide.[6] Crucially, when reward is finally encountered, the values assigned to the earlier state-action pairs that led to it are then updated to reflect the later payoff.[7] Over time, in environments where there are meaningful state-action-reward relationships to be discovered, repeated experimentation with different state-action pairings allows the development of a meaningful representation of value, through these back-tracking computations of the ensuing reward. Effective RL algorithms use representations of value to guide agents to forgo small short-term gains and endure unrewarding actions in order to get to situations where larger rewards can be reaped. A high-value action might itself deliver no reward, but set the agent up so a subsequently available action will be rewarding (such as the position of being able to checkmate in

[5] Because reward in RL is definitive of the agent's goal, it does not quite align with the meaning of "reward" in animal learning experiments, where it typically refers to a substance or event. Unlike RL reward, substances can lose their motivational power: after consuming a large amount of juice, for example, a chimpanzee might no longer aim to get more. To describe the chimpanzee's motivational profile in the more abstract terms of RL, we could instead say that the animal experiences reward when its energy or hydration level approaches a homeostatic set point; perhaps some complex set of such points could operationalize animal well-being (Juechems & Summerfield, 2019).

[6] This representation of value is typically subject to some temporal discounting function to favour reward that comes sooner. Meanwhile, there are many further degrees of freedom in RL theorizing that I am passing over to keep this summary brief. For example, I focus on the value of state-action pairs; some RL value functions are expressed in terms of states. The policy that is used in the value function calculation may be the agent's current policy, or it may be a variant.

[7] Various RL algorithms enable earlier value updating, for example when the reward payoff is still on the horizon, but becoming more predictable. In particular, model-based RL algorithms allow value updates prior to the actual experience of reward, on the basis of a cognitive model that may itself be learned from reward.

two moves, no matter how the opponent plays). The computational back-tracking of assigning high value to actions on paths that end up yielding reward is a species of prediction-error learning; projected reward estimates get revised, over time, to match reality.

This matching between assigned and real value ends up being modally robust, given the character of the training. An RL agent learning to play chess could start with nothing more than the rules of the game and rewards of +1 for a win, zero for a draw, and -1 for a loss; such an agent's initially random valuation of actions might by chance lead to an initial win against a weak adversary. But this valuation will not last on the strength of an accidental success, where the agent would have lost if the adversary had played slightly differently. In the course of training, the RL agent initially relying on the random valuation will soon be pitted against multiple adversaries playing slightly differently, and revise its valuation as it wins and loses. After millions of appropriately varied games, the agent's ranking of the value of moves in different types of situations will generally become reflective of their real tendency to promote victory, not least because the agent's compression of what counts as a relevantly different type of situation will be continually improving. An initially blank-slate agent of this type, AlphaZero, now outperforms prior world champion AI agents trained on expert human game data, just through repeated self-play training (Silver et al., 2018). Chess has a very large state space, so even the best RL agent will not always know what to do, but for many states of the game, AlphaZero can appropriately be described as knowing which move to take. At least in many states later in the game, if AlphaZero ranks one move over others, this valuation is not just correct, but safely correct; in the state AlphaZero currently occupies, and indeed in any close state,[8] the move that AlphaZero represents as best is in fact the best move. In these cases of successful adaptation, the reason why AlphaZero has stabilized on ranking one move higher than the alternatives is that this move actually does lead to the highest reward in relevantly similar predicaments. This safely accurate representation is therefore a factive mental state: factive because its truth is what accounts for its stable existence in the trained agent, and mental because of its role in guiding the trained agent's choice of move.

Good policies favor actions that have shown high value for the relevant situation, exploiting what has been learned to date, together with some exploration of other actions, to test whether undiscovered higher value elsewhere should prompt a further update to the policy.[9] One simple example is the ε-greedy policy, which chooses a random 'exploratory' action some small percentage of the time, and otherwise takes the 'greedy' action currently rated as having highest value, exploiting what has been learned to date. More sophisticated policies gradually explore less when the environment is better known, or in ways that are sensitive to what is at stake.

Any effective policy will ensure that illusions about value naturally tend to be dispelled, over time: states or state-action pairs that are currently assigned high value will be revisited often because of

[8] A close state here is any state that AlphaZero is compressing together with the actual current state (recalling that in large state spaces it is not possible to maintain a separate representation for every possible state). AlphaZero is not just representing states that it has actually experienced; it is generalizing to cover possible states that it could encounter. Note that knowledge of which move is best does not necessarily require that AlphaZero knows the precise value of the move; if one move is much better than the alternatives, then the ordinal judgment that this move is best can be safe even if there is imprecision or slight inaccuracy in its valuations.

[9] For the sake of simplicity, this paper treats exploration and exploitation as distinct classes of action, but some dynamic environments will allow actions to serve both purposes simultaneously, in what has been called 'exploration by exploitation' (Leibo et al., 2019).

that high valuation, but if they end up failing to yield high reward, their valuations will drop. In biological agents, this process is known as extinction: dogs will salivate at the sound of a metronome if this sound has in the past been followed by the provision of meat, but if the meat stops being provided after the sound, the metronome will soon cease to provoke a response (Pavlov, 1927). In this way biological agents are protected from being misled by merely coincidental patterns of reward, and sensitive to changes over time in the reward landscape.

Throughout learning, agents are guided by their current value representations in all but their exploratory actions, but these exploratory actions are crucial to the legitimacy of the agent's value representations. Viewed in isolation, exploratory actions might look like the antithesis of rational agency: they are independent of the agent's best current representation of value, even random. However, as long as the environment is not yet fully known, agents must incorporate some exploratory actions into their overall activity in order to avoid being trapped in local optima while larger reward lies elsewhere: exploration is needed to ensure that the agent's subjective valuations reflect the objective landscape of reward.

The experience of reward precedes learning, in this framework: acting randomly, naïve agents can stumble upon reward without yet having learned anything. Indeed, amnesiac agents who fail to learn anything can go on accruing some ongoing reward by chance. However, agents who learn from their reward-capturing experiences will outperform random agents over time, assuming that there are regularities in the environment of types that the agents can exploit to form suitable action-guiding mental states. The agent with a false or coincidentally true belief can act in a way that happens to yield reward, on the basis of a current pattern of observations aligning with that belief; the agent with knowledge can act in a way that tends to yield reward, on the basis of successful generalizations over its observations, where the truth of these generalizations explains their presence in the agent.

To further characterize the mental states of RL agents, it will help to distinguish two styles of RL. Some agents update their value functions purely on the basis of experienced state-action-reward contingencies; this is model-free RL, the kind of learning that underpins habitual action. However, it is also possible for agents to develop cognitive representations of the relationships between states, creating shortcuts for updating their value functions. In model-based RL, still driven purely by reward, the agent builds a cognitive model of the environment that enables the simulation of never-experienced state-action sequences, simulation that can be used in flexible planning. When the reward landscape changes, a model-free RL agent must experience the new state-action-reward contingencies multiple times to bring its valuations into line with the new reality (through extinction, or through actually stumbling upon new rewards); a model-based RL agent can use its model strategically, to revise valuations for novel state-action pairings in advance of visiting them.

Humans and many other animals, including rats, are capable of both model-based and model-free RL (Daw, Niv, & Dayan, 2005; Dayan, 2009). Because model-free RL caches or stores values for particular state-action pairings, it allows easy but inflexible action when familiar states are visited (driving home on mental autopilot, I take my usual route, switching lanes when triggered by the sight of the exit sign). Model-based RL allows more taxing but flexible on-the-spot updates (having heard on the radio that my exit is closed for emergency bridge repair, I simulate my options using a mental map, and take an earlier exit, perhaps for the first time). Peter Dayan characterizes the contrast between model-free and model-based control as a contrast between imperatives and declaratives. Model-free control is procedural in character—"it specifies directly the choice of action at each state or location as an imperative command"—where model-based control "provides a set of (semantic)

facts about the structure of the environment and the subject in the form of a forward or generative model," where ideally these facts will entail that some particular action is optimal (Dayan, 2009: 214). Model-free RL agents can often know what to do, given their training and what they are now observing; model-based RL agents can also know what is the case in a more detached way.

In model-based RL, successful learning again results in states that are robustly accurate: because incorrect predictions frustrate the pursuit of reward, models tend to improve over time. The idea here is not that everything that the agent incorporates into its model necessarily constitutes knowledge; especially in the early stages of training, the agent can be expected to model various misconceptions about its environment. The idea is rather that knowledge is the natural endpoint of reinforcement learning, the basin of attraction towards which the agent's representations should ultimately converge. Errors in the model will generally drive the agent to behave in ways that tend to correct those errors; safely accurate representations will be stabilized by their propensity to support the capture of reward.[10]

In model-based RL, the agent uses its observations to construct a model of the environment. Observations are a signal of the underlying state of the environment, and RL models extract reward-relevant differences in these underlying states. Every time I drive home, the sensory stream of observations I receive from the various roads I travel will be different, given slight variations in my lane position, traffic, conditions of illumination and weather. For the purposes of efficient navigation, a learned model of the roads will not retain all these variations, but instead distill whatever features make a difference to successful travel. A useful mental map might distinguish the states of being on the main highway and being on the lower-speed side access road, without further distinguishing states involving small differences in cloud cover or illumination. A model whose state space is too fine-grained will tax the memory and processing capacities of the agent, and impede the effective valuation of actions (at the limit, by making the agent take every trip as if it is her first, because the clouds are slightly different this time). A model that is too coarse-grained will bar the agent from taking advantage of some reward-relevant contingencies (for example, if the model only maps spatial relationships, the agent may never learn that an alternative route is faster during rush hour).

With this overview of RL in hand, we can examine the addition of curiosity to RL agents. All RL agents must explore in order to bring their subjective valuations in line with objective reality, and curiosity "aims to provide qualitative guidance for exploration," (Bellemare et al., 2016: 1). With no qualitative guidance at all, uncurious exploration works by periodically inserting completely random actions in the agent's behavior, but this can be inefficient, for example when the random exploratory actions are wasted in probing situations whose outcomes are already known, while important zones

[10] Objection: how can the propensity to gain (future) reward be secured by representations formed by a (past) history of interaction with the environment? One way of handling this problem would be to adopt something like Cameron Buckner's forward-looking theory of content, which draws on the resources of prediction-error correction models of learning such as those in RL. Backward-looking teleosemantic theories such as those of Fred Dretske (1983), and Ruth Millikan (1987) anchor content in the causal history of the individual animal's experiences, or the evolutionary history of the species. Buckner observes that these theories have struggled with explaining how misrepresentation is possible, and with problems of indeterminacy. His forward-looking theory instead takes advantage of the way representations naturally improve over time as we interact with relevant environmental features. In his view, "a representation's forward-looking content (F+) in some environment is thus what it indicates at the limit of its likeliest revision trajectory, given that environment's informational structure" (Buckner, forthcoming b: 18).

further away are left unexplored. Simple count-based methods add reward for visiting rarely visited states, but these are best for small, deterministic domains, with states likely to be visited multiple times. These methods would also require the reward-allocation function to consult memory prior with each exploratory action, to check which options had already been explored, and how often. For agents more like us, better qualitative guidance for exploration is summarized as: "explore what surprises you" (Bellemare et al., 2016: 1). Rather than demanding a memory check, reward can be allocated more simply for the experience of events that violate the agent's expectations. As an illustration of the contrast here, Schmidhuber's television offers a novel, previously unvisited state at every moment we observe it, but it is unsurprising, as we have no expectations concerning the particular configuration of the display at any moment. Stimuli with partial familiarity, on the other hand, set up expectations that can be proven wrong, generating prospects for surprise.

There are various ways to operationalize reward for surprise. Existing methods add reward either for actions that maximize the prediction error of the agent's world model, or for prediction improvement over time[11] (Achiam & Sastry, 2017). Artificial curiosity can combine these approaches (Ten, Kaushik, Oudeyer, & Gottlieb, 2021) or even add a self-model reflecting the agent's awareness of its inner state, enabling a more strategic pursuit of surprise (Haber, Mrowca, Wang, Fei-Fei, & Yamins, 2018). In general, agents who experience surprise as reward do better if they are making predictions involving lasting features of the environment, as opposed to predictions of raw sensory stimulation, for reasons Deepak Pathak and colleagues spell out as follows, in a discussion of RL agents in videogame environments:

> Making predictions in the raw sensory space (…) is undesirable not only because it is hard to predict pixels directly, but also because it is unclear if predicting pixels is even the right objective to optimize. To see why, consider using prediction error in the pixel space as the curiosity reward. Imagine a scenario where the agent is observing the movement of tree leaves in a breeze. Since it is inherently hard to model breeze, it is even harder to predict the pixel location of each leaf. This implies that the pixel prediction error will remain high and the agent will always remain curious about the leaves. But the motion of the leaves is inconsequential to the agent and therefore its continued curiosity about them is undesirable. The underlying problem is that the agent is unaware that some parts of the state space simply cannot be modeled and thus the agent can fall into an artificial curiosity trap and stall its exploration. (Pathak, Agrawal, Efros, & Darrell, 2017: 3)

The pixelated moving leaves represent two problems here: they form patterns too complex for the agent to master, and their motion is an "inconsequential" feature of the environment, so the agent derives no extrinsic reward from standing entranced by it. To protect agents from such fruitless curiosity traps, Pathak and colleagues transform the sensory input in their model into a feature space

[11] It is hard to allocate reward for prediction improvement over time in part because the agent can come to make better predictions simply by switching attention: "the possible naive implementation comparing prediction errors between a window around time t and a window around time $t - \theta$ is in fact nonsense: this may for example attribute a high reward to the transition between a situation in which a robot is trying to predict the movement of a leaf in the wind (very unpredictable) to a situation in which it just stares at a white wall trying to predict whether its color will change (very predictable)." (Oudeyer and Kaplan 2007: 8). To solve this problem, prediction improvement methods need to define a "sensorimotor context" for any given exercise of curiosity; it is suggested that this is not in general a tractable problem (cf. Pathak et al., 2017:3, "there are currently no known computationally feasible mechanisms for measuring learning progress").

just representing aspects of the environment relevant to action. To learn this feature space, a network is trained on the inverse dynamics task which specifies two subsequent states occupied by the agent and asks the network to predict how the agent must have acted to get from the first state to the second (a task that motivates the network to disregard observations irrelevant to action, encoding only features that the agent can control, or features that have an impact on how the agent acts). The next steps introduce a role for surprise: "We then use this feature space to train a forward dynamics model that predicts the feature representation of the next state, given the feature representation of the current state and the action. We provide the prediction error of the forward dynamics model to the agent as an intrinsic reward to encourage its curiosity." (Pathak et al., 2017, 2). This forward dynamics model predicts how things will turn out for the agent as it acts, generating reward for those actions whose outcomes are contrary to its predictions; that is, for surprising outcomes. If curious animals are similarly driven by a mandate to do things whose consequences are unpredictable, this helps to explain the innovative quality of their behavior.

In difficult 3-D video game navigation tasks, Pathak's curious RL agents outperform rivals trained only on extrinsic (game point) reward, while also avoiding nuisance curiosity traps of white noise. In contrast to randomly exploring agents who get stuck in the loops of local minima, these curious agents swiftly learn to navigate hallways systematically, and visit many rooms in large mazes (Pathak et al., 2017: 7-8). Indeed, variations of these curiosity-driven agents can do remarkably well at videogame play across a large range of games, even when learning *purely* on intrinsic reward, with no guidance from game score, doubtless in large measure because games are designed to mesh with natural human mechanisms for intrinsic reward (Burda et al., 2018). However, these agents remain vulnerable to curiosity traps of other types. In particular, if they have the power to generate stochastic events such as coin flips, or if they find a TV whose channel changes randomly in response to some action of theirs, then they can get stuck generating reward for themselves by repeatedly witnessing these mildly surprising actions within their control, ceasing larger exploration or progress within the game.

It is a good question how animals harvest the benefits of curiosity without getting stuck in the traps. Animals do seem to learn from surprise; indeed the standard model of classical conditioning makes the stronger claim that "organisms *only* learn when events violate their expectations" (Rescorla & Wagner, 1972: 75, emphasis added). It is now thought that the phasic dopamine crucial to learning is released not only when animals experience surprising reward, but for surprising experiences of all types (for a review, see Barto, Mirolli, & Baldassarre, 2013). Of course, if surprise functions as reward, reinforcement learning will teach agents to forgo small, short-term surprise in favor of actions with the potential to produce more surprise in the longer term. Interacting with the pens in Hsee's experiment differs from coin flipping in this respect: the coin flip produces a single, unexpected observation of a familiar type, with no consequences for the future, where clicking a yellow-stickered pen might give a novel sensation (so one learns how that feels) while also enabling one to classify the lasting object itself as a shocker or a dud. Having categorized it one way or the other in one's model of the environment creates further expectations that may later be violated, yielding additional possibilities for later surprise (if for example a shocker pen at some point ceases to shock). More generally, stronger long-term expectations can be produced by actions that reveal broader patterns of underlying causal structure; for example, in the patterns of explanation-seeking questioning and causal experimentation that start early in human development (Liquin & Lombrozo, 2020). Theories with greater generality create the possibility for surprising counter-examples, so actions that lead to the formation of such theories will be reinforced.

If surprise functions as reward, state-action pairs will seem good to an agent when similar state-action pairs have in the past led to surprise; more strategically, agents who can anticipate that they don't know what will happen if they act a certain way will be drawn to do so. This is not to say that surprise will always be produced by these actions. Prey-stalking behavior can seem attractive to a predator even if most stalking ends in disappointment, as long as it sometimes pays off; so also, curious behaviors like mailbox-checking can be reinforced by very occasional surprises, even if the most common outcome is a relatively unsurprising empty box. As for avoiding traps, it may help that nonhuman animals typically lack easily available ways of generating stochastic outcomes, and it should be conceded that sometimes humans do get stuck flipping channels or playing mildly surprising but pointless games. Human self-consciousness could also give our curiosity the kind of strategic form that would keep us from dwelling on low-level chancy events. In reinforcement learning, artificial agents can be structured to develop both a world-model (predicting the consequences of the agent's action) and a self-model that tracks the types of situations in which the world model fails; these agents can adversarially choose larger sequences of actions where the world-model can be expected to fail, systematically expanding the boundaries of the agent's knowledge (Haber et al., 2018).

The question of why exactly biological agents need curiosity can now be addressed more directly. Knowledge can be gained even if it is not actively pursued: the basic mechanisms of reinforcement learning work for agents driven strictly by extrinsic reward. AlphaZero is such an agent, with a reward function strictly limited to winning and losing games (+1/-1), and no curiosity whatsoever. In the course of its training, this completely incurious agent nevertheless develops extensive, indeed superhuman, knowledge of how to play chess. A relevant feature of chess is that it is a fully observable environment, with games averaging 40 moves per side. Whenever AlphaZero acts, it is not long before the environment supplies instructive feedback on its success or failure. By contrast, the natural world is very much a partially observable environment, with extrinsic rewards such as nourishment and reproductive opportunities only sparsely distributed among a massive flood of other sensory signals, which in turn provide only limited information about the local causal reality. In such sparse reward scenarios, incurious agents do poorly. So for example, AlphaZero's successor MuZero is superhuman at chess and many Atari games, but humans beat it easily at games of long-range strategic exploration (Schrittwieser et al., 2020, Supplementary Table S1). In the Atari game Montezuma's Revenge, for example, the player must navigate an extended 99-room labyrinth, searching for keys, amulets and other devices to unlock doors and defeat enemies later. The first reward is a key that can be grasped only through navigating a precise path along ladders and ropes, and jumping over a skull; it is estimated that random action sequences will attain this first reward only once every 500,000 attempts (Salimans & Chen, 2018). An average human will score about 4,700 points on this game; incurious MuZero is unable to score a single point after a million rounds of training, with its pure focus on game score. Curious RL does well, however; indeed, artificial agents driven purely by curiosity, with no reward for scoring game points, can do surprisingly well (Burda et al., 2018).

One reason why curiosity is helpful in partially observable environments is that it is up to the exploring agent to decide what to observe, and under the time pressure of a changing world, it is important not to waste time exploring what is already well known. Another reason relevant to biological agents is that curiosity can scaffold the pursuit of extrinsic rewards such as nourishment, whose acquisition in a competitive environment may require extended sequences of action. Gianluca Baldassare describes the challenged posed by sparse extrinsic reward signals as follows: "Learning mechanisms of animals fail to function when there are long delays between the performed

behaviours and the learning signals they cause. Moreover, there are few chances to produce, by trial-and-error, complex behaviours and long action chains that result in a positive impact on homeostatic needs. As a consequence complex behaviours and chains would never be learned based only on extrinsic motivations." (Baldassarre, 2011: 3) The fact that biological learning mechanisms fail after long delays is not something that is easily overcome; we need processes of forgetting and extinction to keep us up to date in a changing environment. However, if curiosity can deliver knowledge of lasting features of the environment that are only indirectly relevant to extrinsic reward, these can function as what Baldassare describes as "readily available building blocks" that can support complex actions later, as needed. In complex environments, animals who experience primary (unlearned) reward only for nourishment and reproductive actions will be outperformed by animals who also experience primary reward for knowledge gain; under the pressure of natural selection, we can expect curiosity to become a favored trait in the primary reward functions (or natural instincts) of animals who may need to learn complex behaviors to satisfy their other needs (Singh, Lewis, Barto, & Sorg, 2010).

Baldassare was the theorist who spoke of the direct detection of levels of knowledge in the brain. He did so in a passage distinguishing extrinsic motivations such as thirst from intrinsic motivations such as curiosity. In his view, extrinsic motivations such as thirst use fitness proxies in the visceral body, which trigger innate and learned behaviours to keep the animal within safe homeostatic limits. For thirst, this role is actually played by neurons in the forebrain, outside the blood-brain barrier, which measure changes in ion concentrations (Leib, Zimmerman, & Knight, 2016). Intrinsic motivations use proxies properly inside the brain itself, such as surprise, in the case of curiosity. In either case, these proximal mechanisms work to secure the distal aim of enhancing fitness. Thirst makes us drink, as curiosity makes us investigate. Cellular function is generally optimized when our fluid balance falls within a narrow range; the adaptive function of thirst is not to produce a certain ion concentration in the forebrain but to maintain the fluid balance of the whole body, which thirst does both by making it pleasurable to drink (positive reinforcement) and by being an aversive experience (negative reinforcement). The proxy can occasionally come apart from the target, for example in thirst-disrupting diseases, but when all is well, the reward signals of thirst ensure that animals can learn behaviors that keep their bodies appropriately hydrated. On the side of curiosity, when the brain detects significant discrepancies between its expected and actual experiences, the resultant feelings of surprise serve as markers of knowledge gain, delivering the pleasure of discovery, while the absence of surprise is generally indicative of a well-known environment. Strictly speaking, what is directly detected in the brain is surprise; however, this signal generally marks a change in our level of knowledge. Again, the proxy is to be distinguished from the target; occasional illusory experiences could be surprising, and could prompt the formation of false beliefs rather than knowledge. The adaptive function of curiosity is not to produce surprise signals in the brain, but to motivate animals to act in ways that accelerate their accumulation of knowledge, an acceleration that seems to be vital for the survival of organisms in challenging environments such as ours.[12]

[12] For discussion and comments on this material, I am grateful to Sara Aronowitz, David Barnett, Nilanjan Das, Jane Friedman, Eliran Haziza, Arturs Logins, Jessica Moss, Daniel Munro, Juan Piñeros-Glasscock, Sergio Tenenbaum, Brian Weatherson, Mason Westfall, Evan Westra, Timothy Williamson, audiences at Princeton University, the University of British Columbia, Lehigh University and Georgia State University. Thanks also to an anonymous reviewer for Oxford University Press for helpful feedback.

References:

Achiam, J., & Sastry, S. (2017). Surprise-based intrinsic motivation for deep reinforcement learning. *arXiv preprint arXiv:1703.01732*.

Aristotle. (1984). *The Complete Works of Aristotle, Revised Oxford Translation* (J. Barnes Ed.). Princeton: Princeton University Press.

Auersperg, A. (2015). Exploration technique and technical innovations in corvids and parrots. In A. B. Kaufman & J. C. Kaufman (Eds.), *Animal creativity and innovation* (pp. 45-72). Holland: Elsevier.

Baldassarre, G. (2011). *What are intrinsic motivations? A biological perspective*. Paper presented at the 2011 IEEE international conference on development and learning (ICDL).

Barto, A., Mirolli, M., & Baldassarre, G. (2013). Novelty or surprise? *Frontiers in psychology*, 907.

Behrens, T. E., Muller, T. H., Whittington, J. C., Mark, S., Baram, A. B., Stachenfeld, K. L., & Kurth-Nelson, Z. (2018). What is a cognitive map? Organizing knowledge for flexible behavior. *Neuron, 100*(2), 490-509.

Bellemare, M., Srinivasan, S., Ostrovski, G., Schaul, T., Saxton, D., & Munos, R. (2016). Unifying count-based exploration and intrinsic motivation. *Advances in neural information processing systems, 29*.

Berlyne, D. E. (1966). Curiosity and Exploration. *Science, 153*(3731), 25-33.

Bringsjord, S., Govindarajulu, N. S., Banerjee, S., & Hummel, J. (2017). *Do Machine-Learning Machines Learn?* Paper presented at the 3rd Conference on" Philosophy and Theory of Artificial Intelligence.

Buckner, C. (forthcoming-a). Black Boxes or Unflattering Mirrors? Comparative Bias in the Science of Machine Behaviour. *British Journal for the Philosophy of Science*.

Buckner, C. (forthcoming-b). A Forward-Looking Theory of Content. *Ergo, an Open Access Journal of Philosophy*.

Burda, Y., Edwards, H., Pathak, D., Storkey, A., Darrell, T., & Efros, A. A. (2018). Large-scale study of curiosity-driven learning. *arXiv preprint arXiv:1808.04355*.

Byrne, R. A., Kuba, M., & Griebel, U. (2002). Lateral asymmetry of eye use in Octopus vulgaris. *Animal Behaviour, 64*(3), 461-468.

Carruthers, P. (2018). Basic questions. *Mind & Language, 33*(2), 130-147.

Carruthers, P. (2023). The Contents and Causes of Curiosity. *British Journal for the Philosophy of Science*.

Dashiell, J. F. (1925). A quantitative demonstration of animal drive. *Journal of Comparative Psychology, 5*(3), 205.

Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature neuroscience, 8*(12), 1704-1711.

Dayan, P. (2009). Goal-directed control and its antipodes. *Neural Networks, 22*(3), 213-219.

Dretske, F. (1983). *Knowledge and the Flow of Information*. Cambridge: MIT Press.

Dubey, R., & Griffiths, T. L. (2020). Reconciling novelty and complexity through a rational analysis of curiosity. *Psychological review, 127*(3), 455.

Friedman, J. (2013). Question-directed attitudes. *Philosophical Perspectives, 27*(1), 145-174.

Glickman, S. E., & Sroges, R. W. (1966). Curiosity in zoo animals. *Behaviour, 26*(1-2), 151-187.

Gottlieb, J., & Oudeyer, P.-Y. (2018). Towards a neuroscience of active sampling and curiosity. *Nature Reviews Neuroscience, 19*(12), 758-770.

Haas, J. (2022). Reinforcement learning: A brief guide for philosophers of mind. *Philosophy Compass*. doi:10.1111/phc3.12865

Haber, N., Mrowca, D., Wang, S., Fei-Fei, L. F., & Yamins, D. L. (2018). Learning to play with intrinsically-motivated, self-aware agents. *Advances in neural information processing systems, 31*.

Harlow, H. F. (1950). Learning and satiation of response in intrinsically motivated complex puzzle performance by monkeys. *Journal of Comparative and Physiological Psychology, 43*(4), 289.

Hsee, C. K., & Ruan, B. (2016). The Pandora Effect: The Power and Peril of Curiosity. *Psychological Science, 27*(5), 659-666.

Juechems, K., & Summerfield, C. (2019). Where does value come from? *Trends in cognitive sciences, 23*(10), 836-850.

Kang, M. J., Hsu, M., Krajbich, I. M., Loewenstein, G., McClure, S. M., Wang, J. T.-y., & Camerer, C. F. (2009). The wick in the candle of learning: Epistemic curiosity activates reward circuitry and enhances memory. *Psychological Science, 20*(8), 963-973.

Kidd, C., & Hayden, B. Y. (2015). The psychology and neuroscience of curiosity. *Neuron, 88*(3), 449-460.

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2012). The Goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. *PloS one, 7*(5), e36399.

Kobayashi, K., Ravaioli, S., Baranès, A., Woodford, M., & Gottlieb, J. (2019). Diverse motives for human curiosity. *Nature Human Behaviour, 3*(6), 587-595.

Kornell, N. (2014). Where is the 'meta' in animal metacognition? *Journal of Comparative Psychology, 128*(2), 143-149.

Kuba, M. J., Byrne, R. A., Meisel, D. V., & Mather, J. A. (2006a). Exploration and habituation in intact free moving Octopus vulgaris.

Kuba, M. J., Byrne, R. A., Meisel, D. V., & Mather, J. A. (2006b). When do octopuses play? Effects of repeated testing, object type, age, and food deprivation on object play in Octopus vulgaris. *Journal of Comparative Psychology, 120*(3), 184.

Lau, J. K. L., Ozono, H., Kuratomi, K., Komiya, A., & Murayama, K. (2020). Shared striatal activity in decisions to satisfy curiosity and hunger at the risk of electric shocks. *Nature Human Behaviour, 4*(5), 531-543.

Leib, D. E., Zimmerman, C. A., & Knight, Z. A. (2016). Thirst. *Current Biology, 26*(24), R1260-R1265.

Lewis, S. A., Negelspach, D. C., Kaladchibachi, S., Cowen, S. L., & Fernandez, F. (2017). Spontaneous alternation: a potential gateway to spatial working memory in Drosophila. *Neurobiology of learning and memory, 142*, 230-235.

Liquin, E. G., & Lombrozo, T. (2020). Explanation-seeking curiosity in childhood. *Current Opinion in Behavioral Sciences, 35*, 14-20.

Loewenstein, G. (1994). The psychology of curiosity: A review and reinterpretation. *Psychological Bulletin, 116*(1), 75.

Millikan, R. G. (1987). *Language, thought, and other biological categories: New foundations for realism*: MIT press.

Montgomery, K. C. (1952). A test of two explanations of spontaneous alternation. *Journal of Comparative and Physiological Psychology, 45*(3), 287.

O'Keefe, J., & Nadel, L. (1978). *The Hippocampus as a Cognitive Map*. Oxford: Clarendon Press.

Oudeyer, P.-Y., & Kaplan, F. (2007). What is intrinsic motivation? A typology of computational approaches. *Frontiers in neurorobotics, 1*, 1-14.

Oudeyer, P.-Y., & Smith, L. B. (2016). How evolution may work through curiosity-driven developmental process. *Topics in Cognitive Science, 8*(2), 492-502.

Pathak, D., Agrawal, P., Efros, A. A., & Darrell, T. (2017). *Curiosity-driven exploration by self-supervised prediction*. Paper presented at the International conference on machine learning.

Pavlov, P. I. (1927). *Conditioned reflexes: an investigation of the physiological activity of the cerebral cortex*. London: Oxford University Press.

Renner, M. J. (1988). Learning during exploration: The role of behavioral topography during exploration in determining subsequent adaptive behavior in the sprague-dawley rat (Rattus norvegicus). *International Journal of Comparative Psychology, 2*(1).

Rescorla, R. A., & Wagner, A. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In *Classical Conditioning II* (pp. 64-99). New York: Appleton-Century-Crofts.

Richter, J. N., Hochner, B., & Kuba, M. J. (2016). Pull or push? Octopuses solve a puzzle problem. *PloS one, 11*(3), e0152048.

Salimans, T., & Chen, R. (2018). Learning montezuma's revenge from a single demonstration. *arXiv preprint arXiv:1812.03381*.

Schmidhuber, J. (2010). Formal theory of creativity, fun, and intrinsic motivation (1990–2010). *IEEE transactions on autonomous mental development, 2*(3), 230-247.

Schrittwieser, J., Antonoglou, I., Hubert, T., Simonyan, K., Sifre, L., Schmitt, S., . . . Graepel, T. (2020). Mastering atari, go, chess and shogi by planning with a learned model. *Nature, 588*(7839), 604-609.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell system technical journal, 27*(3), 379-423.

Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., . . . Graepel, T. (2018). A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. *Science, 362*(6419), 1140-1144.

Silver, D., Singh, S., Precup, D., & Sutton, R. S. (2021). Reward is enough. *Artificial Intelligence, 299*, 103535.

Singh, S., Lewis, R. L., Barto, A. G., & Sorg, J. (2010). Intrinsically motivated reinforcement learning: An evolutionary perspective. *IEEE transactions on autonomous mental development, 2*(2), 70-82.

Small, W. S. (1899). Notes on the psychic development of the young white rat. *The American Journal of Psychology, 11*(1), 80-100.

Sutherland, N. (1957). Spontaneous alternation and stimulus avoidance. *Journal of Comparative and Physiological Psychology, 50*(4), 358.

Sutton, R. S., & Barto, A. G. (2020). *Reinforcement Learning: An Introduction, Second Edition*. Cambridge: MIT Press.

Ten, A., Kaushik, P., Oudeyer, P.-Y., & Gottlieb, J. (2021). Humans monitor learning progress in curiosity-driven exploration. *Nature Communications, 12*(1), 1-10.

Tolman, E. C., & Brunswik, E. (1935). The organism and the causal texture of the environment. *Psychological review, 42*(1), 43.

van Lieshout, L. L., Vandenbroucke, A. R., Müller, N. C., Cools, R., & de Lange, F. P. (2018). Induction and relief of curiosity elicit parietal and frontal activity. *Journal of Neuroscience, 38*(10), 2579-2588.

Wang, M. Z., & Hayden, B. Y. (2019). Monkeys are curious about counterfactual outcomes. *Cognition, 189*, 1-10.

Warden, C. J. (1931). Animal motivation experimental studies on the albino rat. In *Animal Motivation Experimental Studies on the Albino Rat*: Columbia University Press.

Whitcomb, D. (2010). Curiosity was framed. *Philosophy and Phenomenological Research, 81*(3), 664-687.
Whittington, J. C., McCaffary, D., Bakermans, J. J., & Behrens, T. E. (2022). How to build a cognitive map: insights from models of the hippocampal formation. *arXiv preprint arXiv:2202.01682*.
Wikenheiser, A. M., & Schoenbaum, G. (2016). Over the river, through the woods: cognitive maps in the hippocampus and orbitofrontal cortex. *Nature Reviews Neuroscience, 17*(8), 513-523.
Williamson, T. (2000). *Knowledge and its Limits*. New York: Oxford University Press.