The Contents and Causes of Curiosity

Peter Carruthers

Abstract

There has been a flurry of recent work on the cognitive neuroscience of curiosity. But everyone in the field offers definitions of curiosity that are metacognitive in nature. Curiosity is said to be a desire for knowledge, or a motivation to learn about something, and so on. This appears problematic. It either makes it difficult to see how curiosity can properly be attributed to cats and rats (let alone birds and bees), or it commits us to attributing capacities for self-awareness in these creatures for which we lack evidence. The goal of the present article is to offer a re-interpretation of the main findings in the literature: showing how it is possible for creatures to be curious while lacking any conception of their own or others' minds, while at the same time arguing that there is something that a metacognitive conception of curiosity gets right.

1 Curiosity: The State of the Art

Following a period of comparative neglect by psychologists in the twentieth century, the last couple of

decades have seen an increasing body of work by cognitive neuroscientists on the topic of curiosity. (For recent reviews, see Kidd and Hayden [2015]; Bromberg-Martin and Monosov [2020]; Gottlieb *et al.* [2020]; and Lopez Cervera *et al.* [2020].) What we know is that satisfying curiosity, both among humans and other animals, is intrinsically rewarding, thereby serving to reinforce the investigative behaviors that led up to it. And we know that humans and other animals are willing to pay a cost to have their curiosity satisfied. For example, monkeys will give up a significant portion of any reward they might receive simply to know whether or not a reward is coming (Bromberg-Martin and Hikosaka [2009]); and humans will seek to satisfy curiosity about highly negative images and events, even though they know that doing so will be unpleasant (Oosterwijk *et al.* [2020]; Sharot and Sunstein [2020]).

Moreover, although common-sense psychology treats curiosity and interest as separate types of affective / emotional state, it seems that the same reward-networks underlie each (Kidd and Hayden [2015]; Daddaoua *et al.* [2016]; Gottlieb *et al.* [2020]). The difference is just that we describe a creature as 'curious' when it engages in overt investigative behavior of some sort (sniffing, approaching to look closer, or—among humans—asking someone a question), whereas creatures are said to be 'interested' when all that is required is continued attention. For instance, we would describe a monkey that is avidly observing a nearby fight between two males in the troupe as interested in the outcome, whereas when the fight takes place hidden in the nearby forest, a monkey that is motivated enough to walk over to watch would be described as curious. Indeed, even instrumental search and investigative behavior (e.g. when foraging) is increasingly thought to rely on the same or quite similar overlapping processes and networks (Gottlieb *et al.* [2020]), and will be treated alongside curiosity here.[1]

Curiosity and curiosity-like behavior have not only been studied in monkeys, but also in pigeons (Gipson *et al.* [2009]; Fortes *et al.* [2016]) and starlings (Vasconcelos *et al.* [2015]). Moreover, exploratory behavior is extremely widespread among animals that find themselves in novel environments. This should be thought of as a kind of spatial curiosity. For instance, a rat released into a novel maze will spend some time running up and down the various corridors and sniffing in all the nooks and crannies, thereby building up a 'place map' of the layout of the maze in the rat's hippocampus that can be used for navigating thereafter (Wills *et al.* [2010]).

Likewise, honey-bees when they first emerge from the hive spend the first couple of days before they become foragers making exploratory flights around the hive, and they do the same more briefly

[1] Section 4.3 will make a proposal about what differentiates instrumental from intrinsically-motivated informational search.

around any newly discovered nectar source (Wei *et al.* [2002]; Woodgate *et al.* [2016]). Moreover, they, too, thereby construct an allocentric map of the layout of the surrounding environment (Cheeseman *et al.* [2014]). In addition, when a bee is thereafter captured and transported in a black box to a novel location and released, it will initially start out on a straight compass-vector to the hive or to the feeder (depending on where it had been going when captured). When it fails to come across expected landmarks along the route, the bee then breaks off from its straight flight and begins an expanding series of looping exploratory flights until it finds something it can place on its mental map, whereupon it computes a new flight vector to its destination (Menzel *et al.* [2005]).

In the scientific literature, curiosity (and search-behavior more generally) is mostly described in terms suggestive of metacognitive states with meta-representational contents. Curiosity is said to be a desire for knowledge (Litman [2005]; Sharot and Sunstein [2020]), an intrinsic motivation to learn (Gruber *et al.* [2014]), a drive-state for information (Blanchard *et al.* [2015]), or a motivation to reduce the gap between what one knows and what one wishes to know (Loewenstein [1994]). Likewise in the philosophical literature (with a few notable exceptions, discussed below), curiosity is said to be an intrinsic desire for true belief (Foley [1987]; Goldman [1999]), or an intrinsic desire for knowledge (Williamson [2000]). Indeed, such views can seem unavoidable. Realizing that curiosity both causes, and is satisfied by, knowledge acquisition, it is natural assume that it must somehow represent or be about the acquisition of knowledge.

For the most part the scientific literature on curiosity has not drawn attention to the apparent implication that meta-representational capacities are present in mammals, birds, and even bees. But some authors have done so, especially those whose interests are more directly concerned with the distribution of metacognitive capacities among animals and human infants. Thus Goupil *et al.* [2016] argue that since preverbal 12-month-old infants seek information from bystanders when ignorant, they can be said to be aware of their own ignorance, and that their goal is to know the location of the target object. Similarly, Hampton [2001] claims to have discovered that monkeys are aware of their own knowledge and ignorance. (See also: Hampton *et al.* [2004]; Templer and Hampton [2012].) Likewise, Krachun and Call [2009] claim that apes understand what is needed to gain knowledge of the location of a target, since they move to the correct position to be able to see it. And using a similar paradigm, Rosati and Santos [2016] claim to have discovered spontaneous metacognition in monkeys, since the animals look first when they don't know which of two tubes have been bated, but go straight to the correct location when they do. All these findings have been taken to demonstrate nascent forms of self-

awareness of their own mental states in preverbal infants, apes, and monkeys.[2]

If curiosity is a desire for knowledge (or something similar), then we have little option but to claim that all creatures capable of genuine curiosity must have at least a simple understanding of their own minds, in which concepts (or concept-like structures) such as KNOW, IGNORANT, and SEE are embedded.[3] For desires, in general, embed concepts in their contents (Delton and Sell [2014]). In order to want X, one has to have some concept of X (albeit greatly simplified in comparison to related human concepts of X or Xs). In order to want food, one must be capable of identifying (some) foodstuffs; and in order to want knowledge, it seems one must be capable of identifying states of knowledge. Standard accounts of curiosity thus leave one in an uncomfortable position. Either one allows that many creatures (including not just apes and monkeys, but mammals generally, birds, and even bees) have some understanding of their own minds, desiring mental states like knowledge or true beliefs for themselves. Or else one needs to find some way of drawing a principled line, denying that creatures below the line are capable of true curiosity, despite the similarities between their behavior and that of humans, and despite the fact that the underlying brain networks are quite similar to, and perhaps homologs of, those underlying human curiosity.

The goal of the present paper is to show a way out of this conundrum. It will sketch an account of curiosity such that even very simple creatures can be curious without needing to be aware of, nor possess even the simplest theory of, their own minds. But still, since curiosity both causes, and is satisfied by, knowledge acquisition—as well as being caused by one's own ignorance and sustained by learning—it has to be allowed that there is a thin, purely-referential, sense in which curiosity is meta-representational. This will involve finding a place for what might be called 'model-free metacognition'

---

[2] There is also an extensive literature on so-called 'uncertainty monitoring' in monkeys and other animals, which has likewise been claimed to offer evidence of self-awareness in these creatures (Smith *et al.* [2003], [2006], [2014]; Couchman *et al.* [2010]). These findings have recently been critiqued by Nicholson *et al.* [2019], [2021], however, who provide evidence that the sorts of tests of uncertainty conducted with monkeys have little or nothing in common with explicit judgments of uncertainty of the sort generally investigated in humans. In any case, the question of the relationship between uncertainty, on the one hand, and curiosity and interest, on the other, is one that there is no space to pursue here.

[3] The qualification 'or concept-like' is introduced here to preempt criticism from philosophers who insist that concept-possessing creatures must satisfy the full 'Generality Constraint' on concepts (Evans [1982]; Bermúdez [2003]; Camp [2004]). But even if non-human creatures are incapable of freely-recombinatorial thought, many employ symbol-like representational structures with a compositional semantics (Carruthers [2009]).

(so-named by analogy with model-free, or purely habitual or habit-like, decision making; Dickinson and Balleine [1994], [2002]; Gläscher *et al.* [2010]; Dayan and Berridge [2014]). In this model-free sense, even bees can meta-represent their own states of learning and ignorance. But they can do so without employing even the simplest, most nascent, kind of understanding of their own minds.

The sort of model-free metacognition to be defended here should be sharply distinguished from the kind of implicit, purely procedural, metacognition discussed at length by Proust [2014]. For the latter is said to operate without any meta-representational signs or symbols. Rather, metacognition is said to be implicit in the ways in which some cognitive processes respond to and modulate others. Although Proust's [2014] book contains no mention of curiosity, Proust [2019] makes clear that she thinks curiosity, too, should be treated as procedurally metacognitive. The views to be explored and defended in what follows, in contrast, will involve explicit representations (albeit nonconceptual ones) with self-directed meta-representational contents.[4]

2 Curiosity as Questioning

Motivated by concerns like those outlined in Section 1, a small handful of philosophers have argued that curiosity is not a desire for knowledge, nor anything of the sort. Rather, it is an affective—emotional, motivational—state that embeds questions as its contents (Whitcomb [2010]; Friedman [2013]; Carruthers [2018]). Thus 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*. And a bee that flies in widening circles when lost is motivated by a state with the content, *where the hive is*, or perhaps, *what is near here*. These are first-order contents, ones that just require the animal to possess concepts like, WHERE, WHICH, WHEN, and WHAT, as well as concepts for things like food and home. The behavior necessary to answer these questions can either be innately linked to the attitude of curiosity (similar to the link between fear and the fear-face, or between disgust and retching), or it can be acquired through normal forms of reward-based learning (where actions that issue in answers are reinforced); or, more likely, both.

[4] On the other hand, there is some convergence here with the work of Arango-Muñoz [2014] on epistemic feelings. Although Arango-Muñoz does not specifically discuss curiosity, plainly curiosity is an epistemic feeling (an epistemic emotion). And he, too, wants to argue that epistemic feelings represent their objects (in this case, presumably, knowledge-acquisition), but in a way that is accessible to creatures that lack any conception of their own minds. He is not explicit, however, about how such a thing is possible. The current paper can be thought of as attempting to fill that gap.

Friedman [2013] points out that the content of a question is not a proposition; nor do questions have truth conditions. Rather, the content of a question is a set of possible answers. Thus the content of a question like, *where home is*, would be a set of propositions of the form, *home is at place p*. The question is answered when the questioner learns a proposition of that form. So a monkey that is curious about the winner of an ongoing fight is in an affective state with the question as content, *who will win*; and curiosity is satisfied when the monkey learns something with the form, *monkey X won*.

The set of propositional answers that are satisfiers for curiosity presumably need to do more than just match the abstract form of the question. They also need to be formulated in terms accessible to the organism. The proposition, *home is up Main Street and then left on Elm Street* (even if true), is of no use to a bee whose question is, *where home is*—although it might well satisfy a curious human. And conversely, an answer that combines a solar bearing with a measure of distance dependent on flight-speed and visual flow would be of no use to most humans, though it is just what is needed to satisfy the curiosity of a bee. One might suggest, then, that the content of curiosity is a set of question-answering propositions framed in organism-accessible terms.[5]

Carruthers [2018] argues that although the content of curiosity is a set of curiosity-satisfying propositions, the satisfaction-relation itself (namely, learning one of those propositions) is not part of the content. If it were, then curiosity would be metacognitive after all, with learning or knowledge-acquisition being components of its content. He argues that while learning mediates the relationship between curiosity and its content, learning is not itself represented in that content. Hence curiosity can be attributed to bees without conceding that such creatures are capable of meta-representing states of learning or knowledge. The next section will argue, however, that this move may have been too quick, and that there is an important sense in which the content of curiosity is meta-representational after all. But it will turn out that this is not a sense that requires curious creatures to possess any sort of theory or model of their own minds, no matter how simple.

3 The Contents of Questioning

To begin, we take a step back: looking at the kinds of contents possessed by affective states in general, and by desires in particular. We can then return to apply some of the morals to the contents of curiosity.

[5] In addition, as we will see in Section 4.3, what satisfies curiosity is always relative to the animal's goals and background values.

3.1 The contents of desire

There is a familiar distinction drawn by philosophers between the objective and subjective satisfaction of desire. A desire is objectively satisfied if the desired state-of-affairs actually comes about, whether or not the desiring agent ever comes to know of it. In contrast, a desire is only subjectively satisfied if the desiring agent comes to think that the desired state-of-affairs is realized, whether or not it really is. (Many desires, of course, are satisfied in both ways, or in neither.) The distinction has generally been discussed in the context of debates about the nature of human welfare and the good life. Hedonic utilitarians think that what matters is that people should be subjectively satisfied—that they should feel that their desires are satisfied. In contrast, critics like Nozick [1974] and Rachels [1986] have developed arguments designed to show that it is objective achievement that matters, even after one's own death. The debate can be transformed into one about the contents (or satisfaction-conditions) of desire, however, which is how it will be taken here.[6]

Suppose what one wants is that one's grandchildren should grow up to be successful adults. Considered on its face, the content of the desire is, *that my grandchildren grow up to be successful adults*. And then that desire is satisfied—its content is realized—if they do become successful, even if by that time one is long dead. It seems that common sense operates, at least sometimes, with an objectivist construal of the satisfaction-conditions of desire. Someone might say of one long after one dies (having died while one's grandchildren were still young), 'At least he/she got what they wanted: the grandchildren grew up to be successful.' Indeed, people will go to considerable lengths to respect the wishes of the dead, doing things that will objectively satisfy their desires. (This may partly be supported by people's naïve dualism; but even convinced physicalists can make sense of these motivations.) It seems that what someone wants is an objective state of affairs—specified by the proposition in the content-clause of the desire-description—whether or not the person is ever subjectively aware of it.[7]

From the perspective of cognitive science, in contrast, it seems that the relevant notion of desire-content is a subjectivist one. For until one becomes aware that the desired situation obtains one

---

[6] So far as I am aware this is not a question that has been pursued previously in the philosophical literature on desire. On the contrary, discussions of desire routinely take for granted that the content of desire is an objectivist one, as do the papers collected in (Lauria and Deonna [2017]), and as does Schroeder [2020]. Even articles that are focused specifically on the issue of naturalizing the content of desire take for granted that the content of a desire for *p* is the obtaining of the state of affairs *p*, independent of one's awareness of *p* obtaining (e.g. Schulte [2019]).

[7] Unless, of course, part of the content of one's desire is that one become aware of something, like the desire, *to see Paris in the Springtime.* This has a partially subjective satisfaction condition.

will continue to be motivated to do things to achieve it (where possible); and the desire will continue to exist until that happens. Suppose one wants to be rich, and follows the stock markets closely in order to become so. But suppose that, unknown to oneself, there is a very large sum of money languishing in a bank account in one's name, left to one by a rich uncle. Then one is rich, in an objective sense. But, not knowing it, one will keep working the stock market, and keep hoping to become so. Only learning that the desired situation obtains will satisfy the desire in the sense of making it go way. Moreover, it is only awareness of the fulfillment of desire that can provide the reward-signal that strengthens or weakens long-term evaluation of things of that sort. Indeed, only events that have a subjective impact on one can issue in affective learning of the kind that has been heavily studied by cognitive neuroscientists.[8]

Both an end to one's motivation toward action, as well as subsequent evaluative learning, depend on subjective interaction with, or awareness of, the things or events desired, then. But the values the entire system is designed to track are not merely subjective ones. So it would be wrong to embrace a purely subjectivist account of the content of desire, either. On the contrary, the values that desires are designed to track (both innately, and as a result of evaluative learning) ultimately reduce to inclusive fitness, which isn't a mental state, and sometimes isn't realized within one's own lifetime. Hunger, for example (the desire for food), is designed by evolution to insure homeostasis. Merely experiencing oneself as consuming food, or believing that one has eaten food, would not do that. So the best way to characterize the contents of desire, if one is to understand and explain the roles that desires play in the psychology of humans and other animals, requires including both objective and subjective constraints on content.

This conclusion is supported by theoretical accounts of the nature of representation (as employed in cognitive science), such as those offered by Rupert [2018] or Shea [2018].[9] Just as the content of indicative (belief-like) states is some subset of the information carried by that state, so the content of motivational states will be some subset of their normal downstream effects. Which subset? Those that we need to appeal to when explaining how evolution or individual learning has stabilized the

[8] The phrase, 'subjective impact' here is intended to mean something weaker than conscious awareness. For even unconsciously perceived events can issue in affective learning (Kirk-Smith *et al.* [1983]; Greenwald and De Houwer [2017]). Nothing more will be made of this point in what follows, however.

[9] There are, of course, more demanding notions of representation, which have a more restricted range of application. Burge [2010], for example, insists that representation is only present where there are perceptual and other constancies, enabling the organism to represent and navigate an objectively real world. Our interest in the present discussion, however, is in cognitive processes more generally.

functional role of the state in question. Desires are apt to cause real-world effects of the sort specified in their objective satisfaction-conditions; but they are also apt to cause experiences of satisfaction. And it seems that the latter are just as important as the former in explaining the functional role of desire. For it is only the experience of satisfaction that brings motivated action to an end and issues in affective learning.

It seems that from the standpoint of cognitive science, then, learning that the desired state of affairs has occurred, or otherwise experiencing the occurrence of the desired events, should be built into the content of the desire itself. For only then are desires satisfied in the sense that matters for psychological explanation. Might one argue, instead, that the subjective aspect of desire could be built into the attitude (the causal role) of desire, rather than its content? Consideration of the best accounts of intentional content in cognitive science suggests not. For the subjective impact of the desired state of affairs is among the normal effects of the state that we need to appeal to in explaining how the functional role of desire became stabilized; and that means it belongs within the satisfaction-conditions (the content) of desire. In contrast, the functional role of desires—in interacting with beliefs and perceptions to select actions—makes no contribution to the conditions under which a desire counts as satisfied. Rather, it is the contents of those attitudes that fix the nature of the interactions. And as we will see in the discussion that follows in the next section, the subjective satisfaction-conditions of desire play an essential role in those interactions during prospective reasoning and decision making.

3.2 Contents in decision making

One critical role for the notion of the content of a desire is to fit together with the contents of other attitudes to explain and rationalize behavior. And here it might seem that the subjectivist notion of desire-content has no role to play. If what one wants is that the grandkids should be successful, and one comes to believe that paying for their college education will help them to become successful, then that might lead one to set aside funds for such payments. Here the relevant content seems to be, *that the grandkids are successful*, and not, *I become aware of the grandkids being successful*. And the same seems to hold true for the contents of other desires, too. If what one wants is a vacation on Martinique, then coming to believe that there are direct flights from a nearby airport might lead one to access the airline's website to make a booking. Here, again, the relevant content seems to be, *that I have a vacation on Martinique*, and not, *I experience a vacation on Martinique*. So here, too, it appears to be objective satisfaction alone that plays a role in reasoning directed at decision-making.

It may well be that in purely discursive, verbally-based, decision making, only the objective

contents of desire play a role. But this sort of decision making is human-specific (of course), and arguably plays a much smaller role in even human life than philosophers commonly assume. Indeed, there is now a rich empirical literature suggesting that imaginative prospection of future actions and events is our default mode of decision making (Gilbert and Wilson [2007], [2009]; Seligman *et al.* [2013], [2016]; Miloyan and Suddenhorf [2016]). Here one imagines an outcome, or imagines acting to achieve that outcome, or both, and responds affectively to the content of that imagining much as if one were experiencing it (albeit with one's affective response being attenuated in comparison to the real thing). In that case, subjective awareness of the objective fulfilment-conditions of desire is implicit in the perspectival nature of future imagining. In imagining being on vacation on Martinique one visually-represents the imagined banana groves, or olfacation-represents the smell of ocean air, and so on.

It seems that the beliefs that figure explicitly in prospective reasoning are ones that result from imagined simulations of future events. And in that case, they will represent those events as if one were experiencing them. If that is a reliable guide to the contents involved, then it would appear that the desire-contents that these imaginings serve to match will include both objective and subjective satisfaction conditions. What leads one to book a flight to Martinique is not just that the island has banana groves and beaches, but rather the anticipated experiencing of banana groves and beaches. In short, in order to understand the role of desire in prospective reasoning, too, we need to see both objective and subjective conditions of satisfaction as involved.

3.3 The content of curiosity

We can now apply the objective / subjective distinction to the case of curiosity. If one is curious whether one's grandchildren will grow up to be successful adults, then in the objectivist sense the question embedded in one's curiosity is answered positively if they do, negatively if they don't, irrespective of one's own subjective states. (Note that in this objectivist sense, then, one's curiosity / question is satisfied either way, no matter what happens.) But in the subjectivist sense of satisfaction, one's curiosity is only satisfied by learning or coming to believe that one's grandchildren have (or have not) grown up to be successful. Construed in an objectivist manner, there is nothing meta-representational about curiosity. The content of the question is entirely first-order, comprising the states of affairs that would constitute a positive or negative answer. But construed in a subjectivist way, curiosity turns out to have a meta-representational content after all.

In contrast with the case of desire, it seem unlikely that anyone would be prepared to defend a merely-objectivist satisfaction-condition for curiosity. While it is possible to claim that one's desire that

one's grandchildren should turn out successful can be satisfied even if one never knows it, it seems much less plausible to claim that one's curiosity about whether they will be successful can be satisfied by the mere fact that there is an answer out there (and satisfied either way, to boot, whether they are successful or not). Our common-sense perspective is surely that curiosity can only be satisfied by learning, or by some subjective informational change in oneself.

Recall, however, that Carruthers [2018] claims that learning mediates the relation between the content of curiosity (the answer) and its satisfaction. It is via learning the answer that curiosity is satisfied, but learning is not itself part of the answer. Fair enough, it is the proposition learned that satisfies curiosity. But it is unclear that this is more than a verbal maneuver. The content of curiosity is specified by its satisfaction-conditions, surely, just as the content of belief is specified by its truth-conditions. In the latter case, the truth-conditions don't need to mention learning or any other mental state of the believer. The truth-condition for one's belief that one's grandchildren will be successful is just that: one's grandkids being successful. But the satisfaction-condition for curiosity must surely include learning. The mere fact that there is an answer to the curiosity-embedded question out there in the world (one's grandchildren really are successful adults, although there is no evidence of it and one never learns it) surely cannot be enough for curiosity about one's grandchildren to be satisfied.

Consideration of the explanatory role that contents play in cognitive science leads us to a similar conclusion. If one asks what adaptive value curiosity is designed to track, it is surely one that has a subjective component. For the adaptive function of curiosity is to acquire knowledge of the world (or, less commonly, of oneself). Curiosity is only satisfied in a way that is adaptative (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. Moreover, curiosity will continue to motivate action until the appropriate learning takes place (or until the time for learning is past or has become irrelevant). And affective conditioning of successful curiosity-satisfying behavioral strategies can only happen when reward-signals are created by learning the answer.[10]

So from both common-sense and scientific perspectives, it seems the satisfaction-condition for

[10] Here, too, someone might ask whether learning might belong within the attitude (the causal role) of curiosity, rather than its content. But consideration of the explanatory role of content in cognitive science should lead us to reject such a suggestion here as well (as it did in the case of desire). And likewise the contents that fit together with other attitudes in prospective decision making will include a subjectivist satisfaction-condition. When one mentally rehearses actions that might satisfy curiosity, it is the anticipated subjective satisfaction of one's questions that plays the causal / explanatory role.

curiosity is that one comes to know (or at least believe) the curiosity-answering facts. That makes the content of curiosity meta-representational after all. 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. So the content of curiosity does have belief-states or knowledge-states among its satisfaction-conditions. And then our original conundrum returns: how can curiosity and related attitudes be so widespread in the animal kingdom, if curiosity requires those creatures to be capable of meta-representing their own states of mind?

3.4 On sense and reference

It will help to resolve the conundrum if we deploy the familiar distinction between sense and reference, or between mode-of-presentation and worldly truth-maker or satisfier.[11] 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. But the satisfaction of curiosity is not presented to the agent as involving the acquisition of beliefs. Rather, the urge to attend and/or investigate simply disappears once learning happens and curiosity is satisfied. The process that satisfies curiosity involves the acquisition of beliefs, but all that agents themselves need to be aware of are the events or states of affairs that those beliefs are about. Nonhuman agents who are curious, and who have their curiosity subjectively satisfied, need have no conception of belief nor of any belief-like state, no matter how simplified in comparison to the human concept of belief.

An analogy with cases drawn from among mental states that have correctness-conditions, rather than satisfaction-conditions, may help. Consider pain. Although the nature of pain has been heavily debated among philosophers of late, all are agreed in separating its sensory component from its evaluative component (the badness of pain); and although accounts of the evaluative component are hotly debated, most are agreed that the sensory component of pain is a representation with correctness conditions (Cutter and Tye [2011], [2013]; Martínez [2011], [2015]; Bain [2013], [2017]; Barlassina and

[11] The sense / reference distinction is generally associated with the thesis that sense determines, or fixes, reference. That is not part of what is intended here, however. On the contrary, we are assuming a broadly informational or teleosemantic account of reference-determination. It is only the 'mode-of-presentation' aspect of the notion of sense that is intended. The 'sense' of a representational state is the way the world presents itself to the agent when represented in that way. The sense is, as it were, the 'manifest image' (Sellars [1962]) that the world presents to the agent.

Hayward [2019]). What pain sensations represent is that there is a risk of tissue damage (of some magnitude) occurring in a specified bodily location. This is the property that pain-signals have been selected to track, and which explains (in conjunction with the evaluative component) the role that pain sensations have in influencing motivation, cognition, and behavior.

No one thinks, however (nor should they think), that pain sensations are presented to subjects as involving risk of tissue damage. Indeed, most creatures that are capable of feeling pain have no idea what tissues are, nor that they might be getting damaged. So while the reference of pain sensations—their correctness conditions—might include tissue damage, their sense—the way in which they are presented to the agents that have those sensations—is quite different. From the agent's perspective, pains are a sort of ineffable secondary quality of the body. They fit into a quality-space of similarity relations with each other (e.g. stabbing versus throbbing) and with other bodily sensations (e.g. itching), but in themselves, from the perspective of the agent, they just manifest as something like, *that quality there*.

Consider a different example. Feelings of hunger that aren't prompted by external stimuli such as the sight of food are largely caused by signals resulting from a mechanism in the medulla (part of the brainstem), which measures the concentration of glucose in the blood (Rolls [1999]). Simplifying greatly, it seems that the correctness-condition for context-free feelings of hunger is that they should accurately track glucose concentrations. This may be what hunger refers to, or represents. Glucose in the blood is the property that hunger is designed to track, and that (partly) explains the role of hunger in controlling search and consumption behavior. But of course felt hunger does not present itself to subjects *as* a low level of glucose in the blood. This is for a familiar reason: hunger is felt by many organisms, almost all of whom have no idea what glucose is.

The basic idea employed in these examples can be generalized from states with correctness conditions (like pain and felt hunger) to states like curiosity that have satisfaction-conditions. For explanatory purposes, the satisfaction condition for curiosity needs to include learning or knowledge acquisition. Think of this as curiosity's 'referential content.' But the only part of that content that is presented to subjects themselves are the facts learned. Although the satisfaction condition for curiosity about the success of one's grandchildren is that one comes to learn that they are successful (or not), all the agent needs to be aware of is the fact that they are successful (or not). So creatures that lack any conception of knowledge or of learning can be curious and can have their curiosity satisfied, when facts are presented to them (answering the question posed) via learning, but not as learned.

It may be worth noting that the sense / reference distinction cannot be deployed to rescue the

desire-theory of curiosity, according to which curiosity is a desire for knowledge. For if curiosity is construed as a desire, with a proposition as content, then there is no option but to conclude that curious creatures must possess a concept like KNOW or BELIEVE. This is because the desire will have the content, *that I know whether such-and-such*. The metacognitive component can't be dropped from the 'mode of presentation' or 'sense' of this desire, since that would transform it into a desire *for such-and-such*. If one is curious whether one's grandkids will be successful in life, that certainly isn't presented to one as a desire that they be successful. One might also want that; but that isn't what one wants when curious. If one insists on regarding curiosity as a form of desire, then there is no option but to say that what one wants is knowledge, or true belief, or something similar. And that then really does raise a problem for the widespread distribution of curiosity across creatures.

3.5 Purely referential metacognition

It seems that while curiosity and related attitudes do have metacognitive contents, this is a novel sort of metacognition—one that is 'model free' and need not depend on, nor involve, any sort of understanding of minds in general nor of one's own mind, no matter how simple. While the 'referential' satisfaction condition of curiosity includes learning something, or coming to know something, its mode of presentation to the agent comprises just the facts learned or known. So a curious agent needs to be capable of learning, but need have no idea what learning is, nor that it is, itself, engaged in learning.

Consider, then, a bee that has become lost, and is now flying in widening exploratory circles. If we take seriously that the bee is motivated by a sort of spatial curiosity, then it is guided by a state that embeds the question, *where home is* (or something similar). And even if we accept that the satisfaction-condition for curiosity includes learning the answer to the embedded question, this need give us no pause in attributing such states to a bee. For from the bee's perspective, all that happens is this: when it recognizes a familiar landmark it locates itself on its mental map of the local environment, and from that it computes the direction (represented as a solar bearing) and distance (represented in terms of a measure of visual flow) to home. It then breaks off its looping search and flies in the direction indicated. Although it is learning the location of home that satisfies curiosity (and constitutes the satisfaction-condition of its state), the bee itself need not be aware that it has learned anything. Indeed, it doesn't need even the most simplified conception of what learning is, nor need it know anything at all about its own mind.

It seems, then, that we have solved one half of our conundrum: the half that pertains to the content of curiosity. We can distinguish between the (referential) satisfaction-condition of curiosity

(which incorporates learning) and the mode in which the satisfaction of curiosity is presented to the agent (which is an effect of learning). So although the satisfaction-condition of curiosity is metacognitive, states of curiosity can be possessed by creatures that have no awareness of their own minds, and have no conception of what learning is. It remains, then, to inquire about the causes of curiosity. This is where we go next.

4 The Causes of Questioning

Recall that Carruthers [2018] acknowledges that curiosity and related attitudes like interest are caused by one's own state of ignorance, while denying that representations of ignorance need be involved. That claim will be critiqued here.

4.1 Prediction errors

How is curiosity caused? Most likely it begins with a prediction-error (surprise). Something unexpected happening will attract attention. This is not yet curiosity, but more like an orienting response. However, if the item / event is appraised as somehow relevant to the creature's interests or concerns (see Section 4.3), and yet fails to evoke recognition or knowledge of the outcome, then the result will be an affective state with a content such as, *what that is*, or, *what will happen*, which motivates (and provides an urge towards) investigative behavior.

Imagine a city dweller who is walking across a city park, not attending to the scenes around her. (She may be listening to a podcast through headphones.) Nevertheless, even in the absence of focused attention, her visual system will produce a gist-like, summary-statistic, representation of the surroundings (e.g. 'lots of grass, bushes, and trees'; Cohen *et al.* [2016]). But now the visual system detects animate motion in a nearby tree (a squirrel, say). This issues in an error signal, similar to that produced by an 'oddball' stimulus of the sort used in vision labs. That attracts attention. She is now consciously aware of the creature.[12]

So far, nothing metacognitive need be involved. An error signal serves to negate a prior expectation, but it does not represent that an expectation is negated. The sequence is something like this: one is expecting / representing a gist or summary-statistic of purely inanimate vegetation; one then detects (at this stage unconsciously) animate motion, which issues in a signal with the content, *not just*

[12] Note that the 'mismatch negativity response' in the cortex that signals error-detection occurs as little as 100 milliseconds after stimulus onset, fully 300 milliseconds before the stimulus becomes conscious (Dehaene [2014]).

inanimate vegetation. It is the latter that serves to attract attention to the stimulus, resulting in conscious perception of the squirrel.

It seems that most, if not all, instances of curiosity are initiated by error signals in this way. A monkey's curiosity about the winner of a fight between two nearby males will be initiated when her attention is attracted by unexpected loud screams, for example. And a bee's looping search behavior is initiated at some point while flying the compass-bearing that would have taken it back to the hive (had it not been displaced in a black box by an experimenter), when the bee registers the absence of an expected landmark of some sort along the route. If the bee's expectation is that at this point in its flight home it should be passing a free-standing tree, then the error signal will have the content, *no tree* (or some bee-ish equivalent of *tree*). Again, nothing metacognitive need be involved.

4.2 Signals of ignorance

It is one thing for attention to be attracted to something, however, and quite another for it to be sustained. Yet curiosity and interest are sustained affective states. It is an absence of knowledge, or an absence of recognition, that issues in sustained questioning. What keeps the bee flying in widening circles is lack of recognition of any of the landmarks on its mental map of the local environment. What keeps the monkey watching the fight between two males is ignorance of who will win. And what may keep a person attending to the animal whose motion has just been detected in a nearby tree is a failure to recognize it, or ignorance of what it may do next, or something of the sort.

An initial puzzle is how an affective state can be caused and sustained by an absence. How can a lack of knowledge or a lack of recognition cause curiosity? One possibility (the one that will be defended here as most plausible) is that the initial error-signal creates a representation which thereafter competes with nascent representations of object-kinds or event-outcomes. If any of the latter win out, then the question is answered and curiosity is satisfied; but so long as that representation is sustained, attentional and/or investigative behavior continues to be motived (subject to appraisals of continued relevance: see Section 4.3). And since that representation only continues to exist so long as the object fails to be recognized or the outcome fails to be known, it carries the information, *not known*. Then that, combined with its function in sustaining informational search, means that the signal represents ignorance; it has the represented content, *not known*.

In the background of this suggestion is a class of neurally-realistic cognitive models of perceptual and other forms of decision making, which are widely employed across cognitive science. These include drift-diffusion models, leaky competitive accumulator (LCA) models, and others (Usher and McClelland

[2001]; Pleskac and Busermeyer [2010]; Forstmann *et al.* [2016]). What all such models have in common is that they assume neural activity representing the various alternatives (the categories a stimulus might belong to, the actions one is choosing among, and so on) builds up over time at varying rates and with varying degrees of noisy fluctuation in activity; and all now assume that the competing representations are mutually inhibitory (Teodorescu and Usher [2013]). A decision is reached when the first of the competing signals reaches some pre-set criterion, set by the agent in light of a trade-off between speed and reliability, as demanded by the circumstances. (A low criterion will result in a swifter decision, but will be more influenced by noise, and vice versa.)

For example, Dufau *et al.* [2012] employ the LCA framework to model people's performance in a word / not-word task. On each trial a sequence of letters is briefly presented, and participants have to swiftly judge whether they constitute a word or not. With each presentation, activity builds in mutually-competitive fashion for a variety of possible words ('house,' 'louse,' and so on), while at the same time there is a fluctuating level of activity for the not-word option, from which activity in those other alternatives subtracts. If one of the word populations reaches criterion within a fixed time-frame one answers 'word,' whereas if the not-known option fails to drop below a fixed criterion within that time-frame one answers 'not-word.'[13]

This sort of account provides us with a template for explaining how interest and curiosity are sustained. Suppose that a novel mechanical toy is introduced onto the living-room floor. Its unusual sound attracts the attention of the house cat. From that point onward a neural population representing, *not known* competes in a mutually-inhibitory way with neural populations representing various familiar moving things (MOUSE, BALL, PAPER-ON-STRING, and so on). So long as these remain below criterion, and the NOT KNOWN population remains above threshold (and assuming continued relevance: see Section 4.3), the cat will continue to observe the thing, and perhaps engage in overt investigative behavior—approaching and sniffing, patting the toy tentatively with a paw, and so on.

Why should we expect there to be a separate representation for, *not known*, however? Why can't curiosity be sustained merely by the failure of any of the competing kind-representing populations to reach criterion? One answer is that neural competition is ubiquitous in the brain, at all levels of organization (Mysore and Kothari [2020]); and models of neural decision-making routinely assume that there is a separate accumulator for each of the alternatives in question (Usher and McClelland [2001];

[13] Note that the not-word option could equally well be designated the 'not-known' option. For the only basis on which one can judge that something is not a word is by failing to recognize it as a known word.

Pleskac and Busermeyer [2010]; Forstmann *et al.* [2016]). Moreover, mutually-inhibiting competition is especially likely to impact decision making at later stages of processing (Teodorescu and Usher [2013]); and we know, too, that there is always competition among the various motor plans that are active as potential alternative outputs in a given situation (Cisek and Kalaska [2010]).

So consider the cat in our example. 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 behavior in question, and rendering it adaptive, that is what it represents, too. In fact, those motor representations can properly be thought of as what Millikan [1995] calls 'pushmi-pullyu' representations, with both directive and indicative contents.

Alternatively, why should we assign the content, *not known* to the curiosity-sustaining signal, rather than a conjunction of the negated alternatives—*not a mouse and not a ball and not paper-on-string*? The answer, in short, is explanatory generality. It is true that, in each instance of curiosity, the sustaining signal will carry information about a conjunction of negated options. But these will differ from case to case. And as with content-assignment generally (Shea [2018]), it is only the content, *not known* that enables a general explanation of all instances of curiosity via the effects of a common representational cause.

4.3 Appraisals of relevance

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. Consider, again, the woman walking through a park, whose attention is attracted by biological motion in a nearby tree. She fails to recognize the creature (it is a possum, and she has never seen one before). But if she generally takes no interest in the natural world—if she just doesn't care—then she may be incurious, and return her attention to her podcast. In contrast, if she likes to watch nature programs and cares about the diversity of the natural world, she may stop to study the creature and what it is doing.

The proposal made here can draw naturally on the 'saliency' attentional network (better described as a 'relevance' network) that has been characterized in some detail to underlie shifts of

attention to a novel stimulus or to unconsciously-activated memories (Corbetta and Shulman [2002]; Corbetta *et al.* [2008]). This system monitors currently-unattended perceptual representations, as well as currently-unconscious memories and other endogenously-activated representations, appraising them for relevance to the organism's underlying values and current goals. And it can be this same system that helps determine whether to continue attending to something that has attracted one's attention, and contributes to decisions about whether or not to actively investigate by approaching closer, sniffing at the thing, or (in the case of humans) asking a verbal question.[14]

This framework suggests, then, that 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. Where a signal with the content, *not known* is appraised as relevant to something one currently wants or is trying to achieve, the resulting investigative behavior is instrumentally-motivated, whereas when it is appraised as relevant to things one cares about more generally, the result will be intrinsically-motivated, and is a form of interest or curiosity.

4.4. Signals of learning

We noted in Section 1 that satisfying one's curiosity is rewarding, for both humans and animals. It is these rewards that ground affective learning of new investigative behaviors via processes of affective conditioning. And it is an implicit expectation of those rewards that motivates investigative actions. Since the reward-signals in question are caused by (the relevant sort of) learning, it seems that those signals must represent that learning has taken place. At the very least they carry the information that learning has taken place; and their overall role suggests that this amounts to a simple form of representation, of the kind that is routinely appealed to in cognitive science (Rupert [2018]; Shea [2018]). And that means that the signals in question have meta-representational correctness conditions. They will be erroneous if they are generated by anything other than learning, or if they occur when

[14] Although the saliency network has mostly been studied in humans, recent investigations have shown the existence of a homologous or partly-homologous network in monkeys (Touroutoglou *et al.* [2016]) and rats (Tsai *et al.* [2020]). Moreover, given the similarities between mammals and birds in their top-down attentional networks and in the details of their brain-wiring (Mysore and Knudsen [2013]; Sauce *et al.* [2014]; Karten [2015]), it is not unreasonable to postulate that a similar saliency network might exist in birds also. Indeed, all creatures will need to employ some form of mechanism for reorienting attention in the light of background goals and needs, even if that mechanism is only analogous to the mammalian saliency network.

learning has not really taken place.

Reward-signals and affective learning, quite generally, depend on representations of the rewarding event. In connection with receipt of food, drink, money, and many other forms of reward, signals representing the rewarding event are received by regions of orbitofrontal cortex, medial temporal cortex, and the ventral striatum from sensory cortices and the thalamus (Schultz [2000]; Rolls [2014]). There they are matched against stored expected values, and error signals are created accordingly, depending on whether the received value is higher or lower than predicted. One would expect, then, that the reward-value of learning, when curiosity gets satisfied, would have to operate in essentially the same way (albeit implicating distinct or only partially overlapping brain networks). And that means that a signal representing that learning has taken place is received by evaluative systems.

Moreover, many forms of curiosity and interest are sustained, not by any expectation of a discrete future question-answering event ('Which monkey will win the fight?'), but rather by ongoing or incremental learning. A creature exploring a novel environment can be thought of as motivated by the question, *what is around here*. But exploration itself is a process of continual learning, with ongoing rewards being received (albeit diminishing ones, as the local environment becomes increasingly familiar). The creature will continue exploring so long as the activity is sufficiently rewarding; that is, so long as enough learning is still taking place to justify the costs of exploration (see Section 4.5). But that suggests that any learning that happens must generate a signal that is received by evaluative systems to issue in some degree of reward. At any rate, the reward systems need to be sensitive, somehow, to the presence or absence of learning.

It is important to note that signals of learning need only figure in the evaluative processes that occur prior to, and subsequently motivate, ongoing action, as well as assigning value to anticipated and past actions. A creature deciding whether to continue exploring or return to home base, for example, need not have any idea of its own learning. For learning is (like every other kind of affectively-relevant event) evaluated upstream of decision-making processes. The latter processes trade in the common-currency of valence, with actions and outcomes assigned values, which are then integrated with costs and likelihoods to issue in a decision (Levy and Glimcher [2012]; Rolls [2014]; Winstanley and Floresco [2016]; Inzlicht *et al.* [2018]). The result is that a creature's current actions will seem good to it (or not), and it will experience an urge to continue (or not).

It may also be worth noting that since learning can only happen where previously there was ignorance, it is possible that curiosity can be both caused and sustained by just a single signal, rather than two separate ones. This would represent each of the two contents, *not known* and, *now learning*. I

shall not attempt to resolve this here. The important point is that curiosity is, in any case, caused and motivated by signals with meta-representational correctness conditions.

4.5 Model-free metacognition

Curiosity, interest, and other questioning attitudes are kinds of affective state. And like other affective states, they not only result from initial appraisals, but also directly motivate actions of various sorts, depending on the nature of the initial appraisal (LeDoux [2012]). The relevant motor plans (both innate and acquired via previous affective learning) will be activated automatically, and will need to be suppressed in an executive—top-down—manner if they are not to be carried to completion. (Compare the fear-face, the anger-face, and so on.) Since signals of one's own ignorance serve just this one function—entering into appraisals of relevance, which then motivate (or not) investigative behavior of various sorts—there is no need for them to be embedded in anything remotely resembling a theory of one's own mind. They can appropriately be described as 'model free,' and even bees can have signals of ignorance and/or learning that play this sort of role.

It is worth stressing, however, that while the signals of ignorance and learning that initiate and sustain curiosity are model free with respect to one's own mind, they can nevertheless contribute to model-based decision making, where the models in question concern the structure of the environment. Many animals are capable of simple forms of planning, mentally rehearsing some of the actions open to them in various combinations and responding affectively to the results. (For primates see: Mulcahy and Call [2004]; Hanus *et al.* [2011]; Völter and Call [2014]; for birds see: Taylor *et al.* [2010]; von Bayern *et al.* [2018]; Gruber *et al.* [2019].) Any such creature should be capable of flexible anticipation of informational—or learning-caused—reward from various combinations of actions, too. So this would be model-free metacognition encompassed within model-based decision making.

For example, the apes tested by Krachun and Call [2009], who knew that there was food bated in one of three positions, would not have needed to understand anything about vision, visual access, or their own ignorance in order to solve the task. (They may, indeed, have possessed such understanding; but that is not what is at stake here.) Rehearsing various combinations of look-to-see actions they can, as a result of previous affective learning, anticipate which ones will issue in (informational) reward. This need not mean anticipating that those actions will issue in information, or will lead to learning. For this is just anticipatory reward-based decision making. As a result of previous affective learning, the anticipated action seems good to one, that is all. No meta-representations need be involved in the animals' model-based reasoning processes.

Indeed, decisions about whether to sustain or discontinue investigation, too, can result from regular reward-based decision-making processes, integrating expected values with estimates of effort to issue in decisions among options. In fact the same algorithms that animals use when deciding whether to stay or go when foraging can equally well be applied to informational search, computing current costs and informational rewards against the anticipated rewards of directing investigative actions elsewhere (Hills *et al.* [2012]; Kidd and Hayden [2015]).

5 Conclusion

It has been suggested here that curiosity is caused (in part) by a signal that has the correctness condition, *not known* and/or one that has the correctness condition, *learning is happening*. And curiosity itself is an affective state with a question as its content, whose satisfaction-condition includes acquisition of knowledge of the relevant form, serving to answer the question (e.g., *that is an X* for curiosity about what a thing is; *X happened* for curiosity about what will happen next, and so on). But in both respects knowledge and its lack are among the externally characterized correctness and satisfaction-conditions of curiosity. Curious creatures do not need to have any awareness of belief, knowledge, or learning. Nor do they need to have any kind of theory or model of their own minds, no matter how 'pared down' and simplified.

Acknowledgments

I am grateful to Heather Adair, Lea Curtis Fine, Joe Gurrola, Chris Masciari, Shen Pan, Aida Roige, and Moonyoung Song for feed-back on a previous draft; and also for extensive comments received from two anonymous referees for this journal.

Peter Carruthers
Department of Philosophy,
University of Maryland, College Park,
MD 20742, USA.
pcarruth@umd.edu

References

Arango-Muñoz, S. [2014]: 'The Nature of Epistemic Feelings', *Philosophical Psychology*, **27**, pp. 193–211.

Bain, D. [2013]: 'What Makes Pains Unpleasant?' *Philosophical Studies,* **166**, pp. 69–89.

Bain, D. [2017]: 'Evaluativist Accounts of Pain's Unpleasantness', in J. Corns (ed.), *Routledge Handbook of Philosophy of Pain*, Oxford: Routledge Press, pp. 40–50.

Barlassina, L. and Hayward, M. [2019]: 'More of Me! Less of Me!: Reflexive Imperativism About Affective Phenomenal Character', *Mind*, **128**, pp. 1013–1044.

Bermúdez, J. [2003]: *Thinking without Words*, Oxford: Oxford University Press.

Blanchard, T., Hayden, B., and Bromberg-Martin, E. [2015]: 'Orbitofrontal Cortex Uses Distinct Codes for Different Choice Attributes in Decisions Motivated by Curiosity', *Neuron*, **85**, pp. 602–614.

Bromberg-Martin, E. and Hikosaka, O. [2009]: 'Midbrain Dopamine Neurons Signal Preference for Advance Information About Upcoming Rewards', *Neuron*, **63**, pp. 119–126.

Bromberg-Martin, E. and Monosov, I. [2020]: 'Neural Circuitry of Information Seeking', *Current Opinion in Behavioral Sciences*, **35**, pp. 62–70.

Burge, T. [2010]: *Origins of Objectivity,* Oxford: Oxford University Press.

Camp, E. [2004]: 'The Generality Constraint, Nonsense, and Categorical Restrictions', *Philosophical Quarterly*, **54**, pp. 209–231.

Carruthers, P. [2009]: 'Invertebrate Concepts Confront the Generality Constraint (and Win)', In R. Lurz (ed.), *The Philosophy of Animal Minds*, Cambridge: Cambridge University Press, pp. 89–107.

Carruthers, P. [2018]: 'Basic Questions', *Mind and Language*, **22**, pp. 130–147.

Couchman, J., Coutinho, M., Beran, M., and Smith, J.D. [2010]: 'Beyond Stimulus Cues and Reinforcement Signals: A New Approach to Animal Metacognition', *Journal of Comparative Psychology*, **124**, pp. 356–368.

Cheeseman, J., Millar, C., Greggers, U., Lehmann, K., Pawley, M., Gallistel, C., Warman, G., and Menzel, R. [2014]: 'Way-Finding in Displaced Clock-Shifted Bees Proves Bees Use a Cognitive Map', *Proceedings of the National Academy of Sciences*, **111**, pp. 8949–8954.

Cisek, P. and Kalaska, J. [2010]: 'Neural Mechanisms for Interacting with a World Full of Action Choices', *Annual Review of Neuroscience*, **33**, pp. 269–298.

Cohen, M.A., Dennett, D., and Kanwisher, N. [2016]: 'What is the Bandwidth of Perceptual Experience?', *Trends in Cognitive Sciences*, **20**, pp. 324–335.

Corbetta, M. and Shulman, G. [2002]: 'Control of Goal-Directed and Stimulus-Driven Attention in the Brain', *Nature Reviews Neuroscience*, **3**, pp. 201–215.

Corbetta, M., Patel, G., and Shulman, G. [2008]: 'The Reorienting System of the Human Brain: From Environment to Theory of Mind', *Neuron*, **58**, pp. 306–324.

Cutter, B. and Tye, M. [2011]: 'Tracking Representationalism and the Painfulness of Pain', *Philosophical*

Issues, **21**, pp. 90–109.

Cutter, B. and Tye, M. [2014]: 'Pain as Reasons: Why it is Rational to Kill the Messenger', *Philosophical Quarterly*, **64**, pp. 423–433.

Daddaoua, N., Lopes, M., and Gottlieb, J. [2016]: 'Intrinsically Motivated Oculomotor Exploration Guided by Uncertainty Reduction and Conditioned Reinforcement in Non-Human Primates', *Nature Scientific Reports*, **6**:20202.

Dayan, P. and Berridge, K. [2014]: 'Model-Based and Model-Free Pavlovian Reward Learning: Revaluation, Revision, and Revelation', *Cognitive, Affective, and Behavioral Neuroscience*, **14**, pp. 473–492.

Dehaene, S. [2014]: *Consciousness and the Brain*, New York: Viking Press.

Delton, A. and Sell, A. [2014]: 'The Co-Evolution of Concepts and Motivation', *Current Directions in Psychological Science*, **23**, pp. 115–120.

Dickinson, A. and Balleine, B. [1994]: 'Motivational Control of Goal-Directed Action', *Animal Learning and Behavior*, **22**, pp. 1–18.

Dickinson, A. and Balleine, B. [2002]: 'The Role of Learning in the Operation of Motivational Systems', in C.R. Gallistel (ed.), *Stevens Handbook of Experimental Psychology*, New York: John Wiley and Sons, pp. 497–561.

Dufau, S., Grainger, J., and Ziegler, J. [2012]: 'How to Say "No" to a Nonword: A Leaky Competing Accumulator Model of Lexical Decision', *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **38**, pp. 1117–1128.

Evans, G. [1982]: *The Varieties of Reference*, Oxford: Oxford University Press.

Foley, R. [1987]: *The Theory of Epistemic Rationality*, Cambridge, MA: Harvard University Press.

Forstmann, B., Ratcliff, R., and Wagenmakers, E.-J. [2016]: 'Sequential Sampling Models in Cognitive Neuroscience: Advantages, Applications, and Extensions', *Annual Review of Psychology*, **67**, pp. 641–666.

Fortes, I., Vasconcelos, M., and Machado, A. [2016]: 'Testing the Boundaries of "Paradoxical" Predictions: Pigeons Do Disregard Bad News', *Journal of Experimental Psychology: Animal Learning and Cognition*, **42**, pp. 336–346.

Friedman, J. [2013]: 'Question-Directed Attitudes', *Philosophical Perspectives*, **27**, pp. 145–174.

Gilbert, D. and Wilson, T. [2007]: 'Prospection: Experiencing the Future', *Science*, **317**, 1351–1354.

Gilbert, D. and Wilson, T. [2009]: 'Why the Brain Talks to Itself: Sources of Error in Emotional Prediction', *Philosophical Transactions of the Royal Society B*, **364**, pp. 1335–1341.

Gipson, C., Alessandri, J., Miller, H.C., and Zentall, T. [2009]: 'Preference for 50% Reinforcement Over

75% Reinforcement by Pigeons', *Leaning and Behavior*, **37**, pp. 289–298.

Gläscher, J., Daw, N., Dayan, P., and O'Doherty, J. [2010]: 'States Versus Rewards: Dissociable Neural Prediction Error Signals Underlying Model-Based and Model-Free Reinforcement Learning', *Neuron*, **66**, pp. 585–595.

Goldman, A. [1999]: *Knowledge in a Social World*, Oxford: Oxford University Press.

Gottlieb, J., Cohanpour, M., Li, Y., Singletary, N., and Zabeh, E. [2020]: 'Curiosity, Information Demand and Attentional Priority', *Current Opinion in Behavioral Sciences*, **35**, pp. 83–91.

Goupil, L., Romand-Monnier, M., and Kouider, S. [2016]: 'Infants Ask for Help When They Know They Don't Know', *Proceedings of the National Academy of Sciences*, **113**, pp. 3492–3496.

Greenwald, A. and De Houwer, J. [2017]: 'Unconscious Conditioning: Demonstration of Existence and Difference from Conscious Conditioning', *Journal of Experimental Psychology: General*, **146**, pp. 1705–1721.

Gruber, M., Gelman, B., and Ranganath, C. [2014]: 'States of Curiosity Modulate Hippocampus-Dependent Learning Via the Dopaminergic Circuit', *Neuron*, **84**, pp. 486–96.

Gruber, R., Schiestl, M., Boeckle, M., Frohnwieser, A., Miller, R., Gray, R.D., Clayton, N., and Taylor, A.H. [2019]: 'New Caledonian Crows Use Mental Representations to Solve Metatool Problems', *Current Biology*, **29**, pp. 686–692.

Hampton, R. [2001]: 'Rhesus Monkeys Know When They Remember', *Proceedings of the National Academy of Sciences*, **98**, pp. 5359–5362.

Hampton, R. [2005]: 'Can Rhesus Monkeys Discriminate Between Remembering and Forgetting?', in H. Terrace and J. Metcalfe (eds.), *The Missing Link in Cognition*, Oxford: Oxford University Press, pp. 272–295.

Hanus, D., Mendes, N., Tennie, C., and Call, J. [2011]: 'Comparing the Performances of Apes (*Gorilla gorilla, pan troglodytes, Pongo pymaeus*) and Human Children (*Homo sapiens*) in the Floating Peanut Task', *PLoS One*, **6**, e19555.

Hills, T., Jones, M., and Todd, P. [2012]: 'Optimal Foraging in Semantic Memory', *Psychological Review*, **119**, pp. 431–440.

Inzlicht, M., Shenhav, A., and Olivola, C. [2018]: 'The Effort Paradox: Effort is Both Costly and Valued', *Trends in Cognitive Sciences*, **22**, pp. 337–349.

Karten, H. [2015]: 'Vertebrate Brains and Evolutionary Connectomics: On the Origins of the Mammalian "Neocortex"', *Philosophical Transactions of the Royal Society B*, **370**, 20150060.

Kidd, C. and Hayden, B. [2015]: 'The Psychology and Neuroscience of Curiosity', *Neuron*, **88**, pp. 449–460.

Kirk-Smith, M., Van Toller, C., and Dodd, G. [1983]: 'Unconscious Odor Conditioning in Human Subjects', *Biological Psychology*, **17**, pp. 221–231.

Krachun C. and Call, J. [2009]: 'Chimpanzees (*Pan troglodytes*) Know What Can Be Seen From Where', *Animal Cognition*, **12**, pp. 317–331.

Lauria, R. and Deonna, J. (eds.) [2017]: *The Nature of Desire*, Oxford: Oxford University Press.

LeDoux, J. [2012]: 'Rethinking the Emotional Brain', *Neuron*, **73**, pp. 653–676.

Levy, D. and Glimcher, P. [2012]: 'The Root of All Value: A Neural Common Currency for Choice', *Current Opinion in Neurobiology*, **22**, pp. 1027–1038.

Litman, J. [2005]: 'Curiosity and the Pleasures of Learning: Wanting and Liking New Information', *Cognition and Emotion*, **19**, pp. 793–814.

Loewenstein, G. [1994]: 'The Psychology of Curiosity: A Review and Reinterpretation', *Psychological Bulletin*, **116**, pp. 75–98.

Lopez Cervera, R., Wang, M.Z., and Hayden, B. [2020]: 'Systems Neuroscience of Curiosity', *Current Opinion in Behavioral Sciences*, **35**, pp. 48–55.

Martínez, M. [2011]: 'Imperative Content and the Painfulness of Pain', *Phenomenology and the Cognitive Sciences*, **10**, pp. 67–90.

Martínez, M. [2015]: 'Disgusting Smells and Imperativism', *Journal of Consciousness Studies*, **22** (5-6), pp. 191–200.

Menzel, R., Greggers, U., Smith, A., Berger, S., Brandt, R., Brunke, S., Bundrock, G., Hülse, S., Plümpe, T., Schaupp, F., Schüttler, E., Stach, S., Stindt, J., Stollhoff, N., and Watzl, S. [2005]: 'Honey Bees Navigate According to a Map-Like Spatial Memory', *Proceedings of the National Academy of Sciences*, **102**, pp. 3040–3045.

Millikan, R. [1995]: 'Pushmi-Pullyu Representations', *Philosophical Perspectives, 9: AI, Connectionism and Philosophical Psychology*, pp. 185–200.

Miloyan, B. and Suddendorf, T. [2016]: 'Feelings of the Future', *Trends in Cognitive Science*, **19**, 196–200.

Mulcahy, N. and Call, J. [2004]: 'Apes Save Tools for Future Use', *Science*, **312**, pp. 1038–1040.

Mysore, S. and Knudsen, E. [2013]: 'A Shared Inhibitory Circuit for both Exogenous and Endogenous Control of Stimulus Selection', *Nature Neuroscience*, **16**, pp. 473–478.

Nicholson, T., Williams, D.M., Grainger, C., Lind, S., and Carruthers, P. [2019]: 'Relationships Between Implicit and Explicit Uncertainty Monitoring and Mindreading: Evidence from Autism Spectrum Disorder', *Consciousness and Cognition*, **70**, pp. 11–24.

Nicholson, T., Williams, D.M., Grainger, C., Lind, S., and Carruthers, P. [2021]: 'Linking Metacognition and

Mindreading: Evidence from Autism and Dual-Task Investigations', *Journal of Experimental Psychology: General*, **150**, pp. 206–220.

Nozick, R. [1974]: *Anarchy, State, and Utopia*, New York: Basic Books.

Oosterwijk, S., Snoek, L., Tekoppele, J., Engelbert, L., and Scholte, H.S. [2020]: 'Choosing to View Morbid Information Involves Reward Circuitry', *Nature Scientific Reports*, **10**:15291.

Pleskac, T. and Busemeyer, J. [2010]: 'Two-Stage Dynamic Signal Detection: A Theory of Choice, Decision Time, and Confidence', *Psychological Review*, **117**, pp. 864–901.

Proust, J. [2014]: *The Philosophy of Metacognition*, Oxford: Oxford University Press.

Proust, J. [2019]: 'From Comparative Studies to Interdisciplinary Research on Metacognition', *Animal Behavior and Cognition*, **6**, 309–328.

Rachels, J. [1986]: *The End of Life*, New York: Oxford University Press.

Rolls, E. [1999]: *The Brain and Emotion*, Oxford: Oxford University Press.

Rolls, E. [2014]: *Emotion and Decision Making Explained*, Oxford: Oxford University Press.

Rosati, A. and Santos, L. [2016]: 'Spontaneous Metacognition in Rhesus Monkeys', *Psychological Science*, **27**, pp. 1181–1191.

Rupert, R. [2018]: 'Representation and Mental Representation', *Philosophical Explorations*, **21**, pp. 204–225.

Sauce, B., Wass, C., Smith, A., Kwan, S., and Matzel, L. [2014]: 'The External–Internal Loop of Interference: Two Types of Attention and Their Influence on the Learning Abilities of Mice', *Neurobiology of Learning and Memory*, **116**, pp. 181–192.

Schroeder, T. [2020]: 'Desire', in E. Zalta (ed.), *The Stanford Encyclopedia of Philosophy.* <https://plato.stanford.edu/archives/sum2020/entries/desire/>.

Schulte, P. [2019]: 'Naturalizing the Content of Desire', *Philosophical Studies*, **176**, pp. 161–174.

Schultz, W. [2000]: 'Multiple Reward Signals in the Brain', *Nature Reviews Neuroscience*, 1, pp. 199-207.

Seligman, M., Railton, P., Baumeister, R., and Sripada, C. [2013]: 'Navigating into the Future or Driven by the Past', *Perspectives on Psychological Science*, **8**, pp. 119–141.

Seligman, M., Railton, P., Baumeister, R., and Sripada, C. [2016]: *Homo Prospectus*, Oxford: Oxford University Press.

Sellars, W. [1962]: 'Philosophy and the Scientific Image of Man', in R. Colodny (ed.), *Frontiers of Science and Philosophy*, Pittsburgh, PA: University of Pittsburgh Press, pp. 35–78.

Sharot, T. and Sunstein, C. [2020]: 'How People Decide What They Want to Know', *Nature Human Behavior*, **4**, pp. 14–19.

Shea, N. [2018]: *Representation in Cognitive Science*, Oxford: Oxford University Press.

Smith, J.D., Beran, M., Redford, J., and Washburn, D. [2006]: 'Dissociating Uncertainty Responses and Reinforcement Signals in the Comparative Study of Uncertainty Monitoring', *Journal of Experimental Psychology: General*, **135**, pp. 282–297.

Smith, J.D., Couchman, J., and Beran, M. [2014]: 'Animal Metacognition: A Tale of Two Comparative Psychologies', *Journal of Comparative Psychology*, **128**, pp. 115–131.

Smith, J.D., Shields, W., and Washburn, D. [2003]: 'The Comparative Psychology of Uncertainty Monitoring and Meta-Cognition', *Behavioral and Brain Sciences*, **26**, pp. 317–373.

Taylor, A., Elliffe, D., Hunt, G., and Gray, R. [2010]: 'Complex Cognition and Behavioral Innovation in New Caledonian Crows', *Proceedings of the Royal Society B: Biological Sciences*, **277**, pp. 2637–2643.

Templer, V. and Hampton, R. [2012]: 'Rhesus Monkeys (*Macaca Mulatta*) Show Robust Evidence of Memory Awareness Across Multiple Generalization Tests', *Animal Cognition*, **15**, pp. 409–419.

Teodorescu, A. and Usher, M. [2013]: 'Disentangling Decision Models: From Independence to Competition', *Psychological Review*, **120**, pp. 1–38.

Touroutoglou, A., Bliss-Moreau, E., Zhang, J., Mantini, D., Vanduffel, W., Dickerson, B., and Barrett, L.F. [2016]: 'A Ventral Salience Network in the Macaque Brain', *NeuroImage*, **132**, pp. 190–197.

Tsai, P-J., Keeley, R., Carmack, S., Vendruscolo, J., Lu, H., Gu, H., Vendruscolo, L., Koob, G., Lin, C-P., Stein, E., and Yang, Y. [2020]: 'Converging Structural and Functional Evidence for a Rat Salience Network', *Biological Psychiatry*, **88**, pp. 867–878.

Usher, M. and McClelland, J. [2001]: 'The Time Course of Perceptual Choice: The Leaky, Competing Accumulator Model', *Psychological Review*, **108**, pp. 550–592.

Vasconcelos, M., Monteiro, T., and Kacelnik, A. [2015]: 'Irrational Choice and the Value of Information', *Nature Scientific Reports*, **5**:13874.

Völter, C. and Call, J. [2014]: 'Younger Apes and Human Children Plan Their Moves in a Maze Task', *Cognition*, **130**, pp. 186–203.

von Bayern, A., Danel, S., Auersperg, A., Mioduszewska, B., and Kacelnik, A. [2018]: 'Compound Tool Construction by New Caledonian Crows', *Nature Scientific Reports*, **8**, 15676.

Wei, C.A., Rafalko, S., and Dyer, F. [2002]: 'Deciding to Learn: Modulation of Learning Flights in Honeybees, *Apis Mellifera*', *Journal of Comparative Physiology A*, **188**, pp. 725–737.

Whitcomb, D. [2010]: 'Curiosity was Framed', *Philosophy and Phenomenological Research*, **81**, pp. 664–687.

Williamson, T. [2000]: *Knowledge and its Limits*, Oxford: Oxford University Press.

Wills, T., Cacucci, F., Burgess, N., and O'Keefe, J. [2010]: 'Development of the Hippocampal Cognitive

Map in Pre-Weanling Rats', *Science*, **328**, pp. 1573–1576.

Winstanley, C. and Floresco, S. [2016]: 'Deciphering Decision Making: Variation in Animal Models of Effort- and Uncertainty-Based Choice Reveals Distinct Neural Circuitries Underlying Core Cognitive Processes', *Journal of Neuroscience*, **36**, pp. 12069–12079.

Woodgate, J., Makinson, J., Lim, K., Reynolds, A.M., and Chittka, L. [2016]: 'Life-Long Radar Tracking of Bumblebees', *PLoS One*, **11**(8), e0160333.