The Motoric Theory of Curiosity

What is the nature of curiosity? There are two types of account currently in the literature. According to one, curiosity is a metacognitive desire. It is a motivation to acquire knowledge or get true beliefs, for examples. According to the other more recent proposal, curiosity is a desire-like attitude that embeds a question as its content. The present paper proposes a third alternative. It is designed to explain how curiosity might be extremely widespread in the animal kingdom and to better explain how it can admit of degrees (and be satisfied by degrees), as well as to explain how it can be traded off against other values in decision-making. On the proposed account, curiosity directly motivates innate or learned investigative behavior. It makes such behavior seem attractive and renders subsequent learning rewarding. No questions are needed; nor is any contentious form of self-awareness required. The paper begins by critiquing the two existing theories, building on those criticisms to develop the motoric theory thereafter.

Key words: affect; curiosity; ignorance; metacognition; questions; reward

1 The Metacognitive Account

Traditionally, philosophers have analyzed curiosity as a metacognitive desire of some sort (Foley 1987; Goldman 1999; Williamson 2000). A curious person wants *to know* something, or desires *to learn* something, or wants to acquire *true beliefs*. This sort of analysis is generally taken for granted in the empirical literature, too, albeit unreflectively – perhaps merely employing our folk-psychological concept of curiosity (Litman 2005; Gruber et al. 2014; Sharot & Sunstein 2020).[1]

---

[1] Metacognition is cognition that is *about* cognition (Dunlosky & Metcalfe 2009). And for sure curiosity is often *expressed* in metacognitive terms. One says, "I really want to know what happens next," for example. But our topic is the underlying – real – nature of curiosity itself, not our common-sense construal of it. The closest we have been able to find to our own view in the empirical literature is the suggestion by Kidd & Hayden (2015) that curiosity is *a drive-state for information*. It is not entirely clear what they intend, but if by *a drive* they mean a merely-motoric motivating state of a sort that is apt to issue in information, but without containing a representation of information-gain as its goal, then this

Even Loewenstein's (1994) well-known "information gap" theory, which sounds as if it might not require metacognition, is actually stated in metacognitive terms. Curiosity is said to arise from "a discrepancy between what one knows and what one *wishes to know*" (p.93; emphasis added). Such accounts will be referred to collectively going forward as "the metacognitive account." They have rarely been explicitly defended. And where they have been (Gottlieb & Oudeyer 2018) it has been assumed – contentiously – that the states of uncertainty that can ground curiosity are themselves metacognitive in nature, whereas others have argued that this sort of implicit uncertainty is best understood in terms of first-order likelihood estimation (van den Berg et al. 2016a, 2016b; Nicholson et al., 2019, 2021; Khalvati et al. 2021).

The same range of metacognitive analyses can be applied to the emotion of interest.[2] Someone who is interested in something wants *to know* about it, or wants *to learn* what happens next, and so on. Indeed, arguably the only difference between curiosity and interest is that the latter state is ascribed to people when nothing more is demanded of them than continued attention, whereas people are said to be curious in circumstances where discovery requires an overt or covert action of some sort (walking over to look closer, asking someone a question, probing one's memory, and so on). But nothing further will be said about interest in what follows. While the mechanisms underlying curiosity and interest are arguably quite similar (Murayama et al. 2019), it will simplify the discussion to keep just a single focus.

---

might be a nascent version of the motoric theory to be developed in Sections 3 and 4. Our account also has much in common with the theory proposed by Goupil & Proust (2023), who suggest that curiosity results from nonconceptual signals of one's own need for information, resulting in a distinctive sort of metacognitive feeling that drives one's subsequent behavior. We, too, (along with one kind of "questioning" account of curiosity to be discussed in Section 2) rely on the existence of such signals when developing our motoric account in Section 3. But we see no reason to accept Goupil & Proust's claim that metacognitive *feelings* of curiosity play a basic explanatory role (as opposed to being involved in explaining human subjects' overt *reports* of their curiosity). And it seems unlikely that metacognitive feelings are as widely distributed in the animal kingdom as we take curiosity to be.

[2] Not everyone agrees that interest is an emotion, of course (nor that curiosity is, come to that). But these issues aren't relevant for our purposes. For there is no doubt that both are *affective* states, in the broad sense that includes moods, regular emotions, desires, repulsions, and pleasure and pain. That is all we need for our purposes.

There is, of course, an anodyne sense in which curiosity aims at knowledge. For that is its function; and it is the acquisition of knowledge that satisfies curiosity, and brings it to an end. But it doesn't follow that a curious creature *represents* the state of knowing as its goal. In the same way, hunger aims at the consumption of nutrients (for that is its function, and it is the consumption of nutrients that makes hunger go away). But a hungry creature need not represent *consuming nutrients* as its goal. It may simply aim at *eating that*, or may just have a motor urge to consume the thing in question. In what follows we assume that all three of the accounts of curiosity we will be discussing are to be construed realistically, as theories of the representations and attitudes that figure in the mind of a curious creature. So we take the metacognitive account, in particular, to claim that curiosity is an attitude (a functional desire-like state) that embeds a representation like KNOW, or LEARN, or something of the sort.[3]

One final clarification: as with other affective states, curiosity admits of both occurrent and dispositional varieties. Someone can feel fear when seeing a dog loose in the park but can also, more generally, be afraid of dogs. Indeed, even while asleep it can truly be said of the person that he fears dogs. Presumably this means, "is disposed to feel fear when meeting a dog," or something of the sort. Likewise, someone can feel curious which team won a recent soccer match when reminded that it is now finished; but in a dispositional sense they might have been curious about the outcome all along. Our focus, here, will be on occurrent, active, episodes of curiosity.

The main problem faced by metacognitive accounts of curiosity is that they are conceptually quite demanding. This makes it hard to see how many types of creature besides ourselves could be curious. For curiosity-like behavior, at least, is widely shared across mammals and birds. Many animals will approach, sniff, examine, and otherwise investigate a novel object. And both monkeys and pigeons (as well as humans) will pay a cost simply to know whether or not a reward will arrive in the future (Bromberg-Martin & Hikosaka 2009;

---

[3] As is usual in the much of the literature, we here use small capitals to designate mental representations, where representations are understood to be symbolic structures of some sort (perhaps realized in rates of firing of particular groups of neurons). And by an *explicit* representation, we do not mean one that is conscious, just that it involves a symbol of some kind that is *about* the content of the representation (in this case, *knowledge*). We should emphasize, however, that explicit representations can include both conceptual and nonconceptual (analog-magnitude, graded) forms of symbolic mental structure.

Vasconcellos et al. 2015; Bennett et al. 2016; Fortes et al. 2016) – and note that this is in circumstances where there is nothing the animal can do to influence the likelihood of the reward itself. Moreover, when infant rats are placed in a novel maze they run up and down the corridors, sniffing in all the nooks and crannies, thereby building up a place-map in their hippocampus corresponding to the spatial layout of their environment (Wills et al. 2011).

Indeed, information is a valuable commodity, and it seems that almost all living creatures need it, not just mammals and birds. Sometimes creatures might engage in a random walk through the environment, scooping up information as they go, attending to and learning from anything novel (Baranes et al. 2014; Poli et al. 2022). But much information-acquisition results from motivated and targeted searches. For organisms tend to need not just any-old kind of information, but rather information relevant to their needs and life-ways. Even bees engage in exploratory flights around the hive before they become foragers, from which they (like rats) construct a crude mental map of the surrounding landscape for use when foraging thereafter (Menzel at al. 2005; Cheeseman et al. 2014). And honey-bees closely observe (exhibiting curiosity-like behavior) the figure-of-eight dances of other bees, from which they learn the distance and direction of the substance the dancer-bee carries (Seeley 1995). Moreover, bumblebees will visually observe the successful actions of other bees, thereby acquiring novel problem-solving behavior (Bridges et al. 2023).

On a metacognitive account, the weanling rats-pups studied by Wills et al. (2011) must want *to know what is around here*; the pigeons studied by Gipson et al. (2009) must be wanting *to learn whether food will arrive*; and the bumblebees studied by Bridges et al. (2023) must be wanting *to know how that thing opens* – or some or other variation on such metacognitive desires. But it is controversial whether even monkeys are capable of representing their own or others' states of mind as such (Penn et al. 2008; Le Pelley 2012; Heyes 2015). So we are forced either to deny that rats, pigeons, and bees can be genuinely curious (postulating some other kind of motivating state to explain their behavior), or to take on theoretical commitments for which we lack independent evidence.

We know of no data suggesting that the mechanisms underlying curiosity-like behavior in insects are at all similar to those that motivate such behavior among mammals. But as noted above, we do at least know that birds (Vasconcelos et al. 2015; Fortes et al. 2016) and monkeys (Bromberg-Martin & Hikosaka 2009), just like humans (Bennett et al. 2016), will pay a cost to acquire information about upcoming rewards, even though that information serves no instrumental purpose. Moreover, we know that among primates, at least, the same neural mechanisms underlie the intrinsic reward-value that attaches to novel information as well as underlying rewards of other sorts (Bromberg-Martin & Hikosaka, 2009, 2011; Blanchard et al. 2015; Daddaoua et al. 2016; Wang & Hayden 2019). So we can be confident that genuine curiosity extends well beyond creatures that are known to be capable of representing their own states of mind.

We can conclude that metacognitive accounts of curiosity are problematic (at least when construed realistically, as requiring some form of mentalizing capacity that enables the creature to represent states of knowledge or learning). This motivates a search for a replacement. We next consider questioning accounts, which have been designed, in part, to overcome this difficulty.

2 The Questioning Account

Contrasting with the metacognitive account, some philosophers have recently proposed an alternative, according to which curiosity is a first-order (non-metacognitive) motivational state that embeds a question as its content (Whitcomb 2010; Friedman 2013; Carruthers 2018, 2021). To be curious about what will happen next, on this view, is to be in a desire-like state with a question as content – *what happens next* – where the attitude taken towards that content motivates actions that are designed to secure answers to the question. Likewise, to be curious where something is, is to be in a desire-like state with the content, *where it is*. To be curious when food will arrive is to be in a desire-like state with the content, *when food will arrive*. To be curious about something's identity, is to be in a state with the content, *what that is*. And to be curious who will win a fight or competition is to be in a motivating state with the content, *who will win*. And so on.

One of the main arguments offered in support of the questioning account of curiosity is that it is conceptually less demanding than the metacognitive account, making it possible to see how many types of creature besides ourselves can be curious. For it requires only simple concepts (or concept-*like* representations, if one embraces especially demanding standards for genuine concept possession, as some philosophers do; Evans 1982; McDowell 1994), such as WHAT, WHERE, WHEN, WHO, and perhaps some others (such as FOOD, if a pigeon can be curious *when food will arrive*). Even these demands are not trivial, however. For it is one thing to be capable of representing specific times and temporal intervals together with particular spatial layouts and directions (which we know that even insects can do), and quite another to be able to represent the general properties involved, as do the concepts WHEN and WHERE. It would be an advantage if we could devise an otherwise-successful account that makes even fewer cognitive demands than this (as Section 3 will show the motoric account can do).

What is also problematic for the questioning account is that much learning takes place incrementally, by degrees. Exploring bees or rats will gradually become acquainted with their surroundings, and a curious infant will become increasingly familiar with a new mechanical toy. So curiosity can be progressively reduced and gradually fade. But it seems that a question has either been answered or it has not. Of course, multipart questions can admit of partial answers. But it is unclear that simple questions do. If one asks, "What is near here?" (as, by hypothesis, a rat-pup placed in a novel maze will do) and someone answers, "The Adventist church is a block in that direction," then one's question has been answered (either helpfully or not, depending on one's purposes). So it seems that any amount of spatial learning, no matter how small, would fully answer a spatial question. But that is not what we observe in the rat-pup's behavior.

In response, it can be said that while the content of a verbal question (its literal meaning) is comprised of a set of possible answers (Karttunen 1977), what speakers intend or mean on particular occasions of use can be graded and more specific. If a visitor to an unfamiliar city has just agreed with a friend who lives there that it is time to go for lunch, then only mentions of restaurants, cafes, and so on will count as answering what the speaker means by asking, "What is near here?" – and a reply that mentions the Adventist church will not qualify. If one is a prospective purchaser being shown around a home, in contrast, then mention of nearby churches could well count as part of an answer. Notice, moreover, that a *full* answer would mention everything that the questioner would consider relevant, including perhaps restaurants, cafes, playgrounds, churches, and bus-stops, depending on the context. So it does seem that answers to what a speaker intends by a question can come in degrees. And if the speaker-meaning of a question is the content of the speaker's underlying questioning attitude, then curiosity can come in degrees, too.

To accommodate incremental satisfaction of curiosity within a questioning account, then, it seems that the contents of the alleged questioning attitudes will need to be enriched. Instead of being motivated by a simple desire-like attitude with the content, *what is around here*, exploring animals will need to be motived by a question such as, *what the main landmarks are around here more-or-less sufficient for navigating thereafter*, or something of the sort. Answers to graded questions like this could plausibly come in degrees. But it is intrinsically less likely that bees (or even rats) might entertain representations with this degree of complexity and sophistication. Of course there must be *something* in the mind of the bee or the rat that gradually reduces curiosity-like motivation with learning; but it seems unlikely that it should be an explicit (symbolically-represented) question with this sort of complexity.

It might be advantageous, then, if an account of curiosity could be developed in terms of representations that are themselves graded in nature (specifically, nonconceptual or analog-magnitude ones), permitting conceptual contents but not requiring them. This would provide maximal coverage of the animal kingdom, enabling us to side-step issues to do with concept-possession. And it might allow us to account smoothly for cases of gradual satisfaction of curiosity and gradual loss of interest, as well. That is one goal of the account to be offered in Section 3. Moreover, even if one is unconvinced by the current critique of the questioning theory of curiosity, it might still be instructive to build the simplest model of curiosity one can, to be evaluated alongside the others. This is because (other things being equal) simpler theories are generally preferable.

A different sort of worry about the questioning account is that it is in danger of collapsing back into a metacognitive one. For all affective states are initiated via an appraisal of some thing, fact, or event. Fear results from appraising a thing or situation as dangerous, disgust results from appraising something as a potential poison or contaminant, and so on. But what sort of appraisal could prompt a desire-like questioning attitude except an appraisal of one's own ignorance? If so, then that would then seem to make the appraisal process itself metacognitive in nature, thereby undercutting much of the rationale for adopting a questioning account in the first place.

Carruthers (2018) responds to this difficulty by pointing out that causation by ignorance is not the same as representing one's own ignorance. So there need not be any metacognitive representations involved in the initiating appraisal process. This certainly seems possible. If considered from the perspective of widely-accepted neural-accumulator models of recognition and judgment, for example, then curiosity might be initiated whenever the accumulators for things or events that are deemed relevant fail to reach criterion. (We will return to the question of how relevance is determined in the appraisal processes that issue in curiosity in Section 4.) The only representations involved could be (failed) representations of things or outcomes.

To elaborate, models of the type appealed to here include drift-diffusion models, leaky competitive accumulator (LCA) models, and others (Usher & McClelland 2001; Pleskac & Busermeyer 2010; Forstmann et al. 2016). What they all have in common is the assumption that neural activity representing the various alternatives (the categories that a stimulus might belong to, the possible outcomes of an event one is observing, and so on) builds up over time at varying rates in response to sensory evidence, with varying degrees of noisy fluctuation; and all now assume that the competing representations are mutually inhibitory (Teodorescu & Usher 2013). Classification occurs, or a specific prediction is formed, when one of the competing signals is the first to reach some pre-set criterion, set implicitly by the agent in light of a trade-off between speed and reliability, as demanded by the circumstances.

For example, consider a female chimpanzee who notices the alpha male issuing a fierce threat-grunt to an adolescent male foraging nearby. Neural populations representing *the alpha will win* and *the adolescent will win* become active, mutually inhibiting one another. In light of the female's previous experience, the former population swiftly reaches criterion (meaning she has now formed a determinate prediction that the alpha will win), and so she loses interest and pays no further attention. But now consider a case where she hears the beta-male of the troupe issuing a threat-grunt to the alpha. As before, two neural populations become active, representing *the alpha will win* and *the beta will win*. But now, in light of her background knowledge, the two populations inhibit one another sufficiently strongly that neither reaches criterion (in effect, implicitly representing that she is ignorant of the outcome). This (in light of its relevance to her own future social behavior) causes her to form a desire-like attitude with the content, *who will win*, which then motivates her to look up and watch. She is now curious about the outcome.

More recently, Carruthers (2021) has developed a rather different (and we think better) response to the objection. He argues that if we assume that *competing* accumulators are the norm, then an optimal trade-off between speed and reliability might be achieved by having a separate neural accumulator that signals ignorance. This would become activated whenever others are also active that pass the relevance test. This accumulator would compete with those representing familiar categories or outcomes, and would result in curiosity if it reaches criterion first. In support of this suggestion Carruthers appeals to word / not-word tasks, in which people are required to respond as quickly and accurately as possible whether a presented sequence of letters is a word or not. These tasks bear a close resemblance to knowledge / ignorance tasks, and have been successfully modelled using a separate accumulator that builds its activity in response to failures of recognition (Dufau et al. 2012). Moreover, he points out that if curiosity were to be initiated merely by failure of category-specific or outcome-specific accumulators to reach criterion, then curiosity could not start until some sort of time-limit on failure had been reached. Yet the onset of curiosity is often almost immediate.

Since the accumulator we are postulating carries information about ignorance as part of its sustaining function, the neural population in question *represents* ignorance (Rupert 2018; Shea 2018). This would make the appraisal process that issues in curiosity metacognitive, of course. However, Carruthers (2021) argues that it is innocently so. For it can be a sub-personal signal with a metacognitive correctness-condition that plays just this one role. It represents ignorance directly (*de re* in philosopher's parlance) rather than as such, and so need not contribute to self-awareness or any other contentious form of metacognition. In fact Carruthers describes it as a form of "model free" metacognition, which is independent of even the very simplest model of one's own or others' minds, and so could well exist in quite simple creatures.

It seems that the questioning account of curiosity can successfully offer a story of one or other of the above sorts about the initial appraisal processes without being self-undermining – that is, without attributing contentious forms of metacognitive awareness to the very creatures it was (in part) designed to avoid. So the objection that the questioning account is in danger of collapsing back into a metacognitive one has been answered.

Nevertheless, postulating that curiosity embeds a question as its content marks this state as unique among affective states generally. In all other cases, appraisal of an indicative representation of an object or event issues in a valenced representation of that very thing. Appraising the cake as delicious makes it (the cake) seem good and attractive, and causes an impulse to obtain it; appraising the grizzly-bear as dangerous makes its approach seem bad and to be avoided, while also producing an impulse to run or freeze, depending on the circumstances; appraising a remark as an insult issues in anger directed at what was said; and so on. In the case of curiosity, in contrast, an appraisal of ignorance is said to issue in an entirely different sort of representation: a question directed at the topic one is ignorant of. Moreover, it is unclear what it could mean to say that the question itself (its content) is valenced (seen as intrinsically good or bad).[4]

This might not be a decisive objection. For a question-theorist can of course allow that there can be valence *associated with* a state of curiosity, albeit not attaching to the content itself. Frustrated curiosity will be experienced negatively, just like any other frustrated desire, whereas satisfied curiosity will be rewarding. And if one has to suppress the motor-urges involved, then that will be experienced negatively, too, since executive control in general is effortful (negatively valenced; Kool & Botvinick 2013).

All the same, if no valence attaches to the content of curiosity itself, then it is unclear how curiosity could figure in decision-making processes involving other types of affective attitude. Many of the psychologists who work in the fields of prospective decision-making (Gilbert & Wilson 2007, 2009), neuroeconomics (Levy & Glimcher 2012; Ruff & Fehr 2014; Sescousse et al. 2015), or animal foraging and decision-making (McNamara & Houston 1986; Shizgal & Conover 1996; Montague & Berns 2002; Mishra 2014) regard valence as the common-currency of value in the kinds of decision-making they study. And indeed, even bees can adaptively trade off variations in the strength of sucrose solutions against noxious degrees of heat (Gibbons et al. 2022). Many assume, too, that decision-making computes over values and likelihoods to issue in a representation of expected value, where valence is a *representation* of the current, contextually determined, degree of value attaching to the object of appraisal.

Indeed, Carruthers himself (2024) argues for just such a position, drawing on the account of representation in cognitive science provided by Shea (2018). For example, when deciding between a proffered food item and a risky monetary gamble, it seems that people implicitly compute the expected value of each choice by integrating outcome values (signaled by degrees of anticipatory valence) with estimates of likelihood (Chib et al. 2009). And we know that curiosity, too, can be traded off against other values, as when one is both curious what the bear will do and fearful of its approach. But it is unclear how such comparisons could be made if the content of curiosity itself (in this case, *what the bear will do*) is not valenced.

---

[4] Questions can be good or bad in quite a different sense, of course. One question can be better than another because it is more likely to issue in knowledge, insight, or understanding. But if some degree of positive valence is to be attached to a question on this sort of basis, then questioners would first have to appraise the likely value of the answers that might be received. And this really would return us to a form of metacognitive account.

The only obvious option is to say that deciding whether to act on one's curiosity (standing one's ground to observe the bear, rather than fleeing) requires one to anticipate how rewarding it would be to get an *answer* to one's question (thereby learning something). But now it really does look as though we might have returned to a richly metacognitive account. This, combined with the problems the questioning account has in dealing with the graded nature of curiosity and its likely extent across the animal kingdom, provides us with good reason to look for an alternative.

3 The Motoric Account

As noted earlier, all affective states are initiated by an appraisal of the personal significance of some object, event, or situation, where the representations that figure in the appraisal process can be either conceptual (such as recognizing an approaching animal as a grizzly-bear) or nonconceptual (such as a sudden loud growl emanating from a nearby bush). And all issue in automatic activation of innate or learned motor plans, whether expressive (the fear face, the anger face, the disgust retch, and so on) or instrumental (an impulse to run from the grizzly, an impulse to withdraw from contact with animal feces). These motor impulses are directly and swiftly produced by the initial appraisal process, independently of any kind of decision-making, and they need to be inhibited if they are not to be carried through to completion (Frijda 2010; LeDoux 2012; Adolphs & Anderson 2018). Moreover, as also noted previously, it seems that all affective states are valenced, making the appraised object or event appear to some degree good or bad. (The approaching grizzly seems bad; the slice of chocolate cake seems good.)[5]

Desires, too, initiate automatic approach motor-impulses (or avoidance ones, in the case of negative desires or repulsions). The same is true of any value-laden stimulus. In a study by Krieglmeyer et al. (2010), for example, participants used a keyboard to move an avatar on the screen upwards or downwards. (They were told to imagine that they were the avatar in question.) The avatar could appear in the top or bottom portion of the screen, and on each trial a positive or negative word was displayed in the center. They were told to move the avatar in one direction if

---

[5] Most affective states also issue in some degree of arousal or de-arousal, including changes in heart-rate and breathing rate, as well as release of chemicals like adrenaline and cortisol into the blood-stream. But this aspect of affective states, although real and important, will play no part in the discussion that follows.

the word was a noun and in the other if it was an adjective. (The valence of the word was irrelevant to the task and could in principle be ignored.) Participants were faster to move the avatar towards a positive word and away from a negative one, no matter whether the avatar was moving upwards or downwards. It seems that the automatic and goal-independent motor-instructions following an appraisal of positive or negative value are quite abstract – "get closer to that thing" versus "get away from that thing" – while the movements one actually performs to implement those instructions can be very different (pressing one key rather than another to make the avatar move upwards or downwards).

The motoric theory of curiosity builds on these general features of affective states, assuming that some set of species-specific action-tendencies will automatically be activated by the initial appraisal process that issues in curiosity. These will include (across all creatures) tendencies to pay attention to the object of curiosity, combined with species-specific dispositions to sniff, approach, observe from other perspectives, and/or manually explore or contact the object in question.

As for the initiating appraisal process itself, we can co-opt the proposal made by Carruthers (2021), but thereafter dropping questions out of the picture altogether (at least when explicitly or symbolically represented, rather than being left implicit in a procedure of some sort; see our later discussion). As previously, model-free signals of ignorance will issue in automatic motor-activation in advance of any sort of decision-making. Tendencies to attend, to approach, to sniff, to look closely, and so forth will become active, depending on the nature of one's ignorance and the affordances of the context, and will issue in such actions unless inhibited. At the same time, representations of those anticipated actions and their likely consequences will be appraised based on one's innate values or previous reward-history, issuing in positive or negative valence attaching to those representations themselves. So approaching the unknown object will seem good and attractive, as will attending to an ongoing event like a fight between two group members. And although the latter might sound metacognitive ("attending seems good"), it need not be so. Positive valence attaching to the motor-urge that controls attention (or motor-*like*, in the case of covert attention) need not involve any representation of attention as such, but can still be traded off against other values. These valences can then figure in standard common-currency computations – implicitly trading off the value of learning against other affective goals that are active in the context to issue in a decision.

Consistent with this suggestion, we know that sensory “forward models” are automatically produced from efference copies of any set of motor instructions, and are used for fine-tuning and correcting subsequent action-execution (Wolpert & Kawato 1998; Wolpert & Ghahramani 2000; Jeannerod 2006). The motor instructions for a right-handed reach, for example, will automatically issue in fine-grained representations of what the subsequent movements should look like and feel like, which can be matched against one’s ongoing experience. And there is reason to think that these sensory forward models are almost ubiquitous in the animal kingdom, even being employed by dragonflies (Mischiati et al. 2014). But there is also good reason to think that forward models – representations of future or potential actions – are routinely evaluated and assigned valences. For all action involves choice: choice between acting and not acting, as well as selection among different ways of implementing those actions. These are evaluated for their energetic and other costs, and are assigned degrees of negative valence accordingly (that is, estimates of how effortful they are; Cisek & Kalaska 2010; Cos et al. 2014; McNamee & Wolpert 2019).

We just have to suppose, then (consistent with the literature on prospection; Gilbert & Wilson 2007, 2009; Seligman et al. 2016), that forward models of automatically-activated investigative actions are not only evaluated for their biomechanical costs but also for their overall anticipated reward-value. (The rewards in question can be either *primary* / innate or *secondary* / learned.) Innately-activated motor outputs of curiosity (attending, approaching, sniffing, and so on) might come with a default positive value, which can then be adjusted upwards or downwards through standard forms of conditioning and evaluative learning. And since all theories of curiosity agree that information-acquisition is experienced as rewarding (with the extent of the reward being a function of the amount learned as well as its relevance to the concerns that prompted one’s curiosity), then novel forms of behavior can become part of the automatic motor output of states of curiosity based on one’s previous reward history. So human children can rapidly learn to ask verbal questions when curious, as will be discussed in Section 4.

The motoric theory seems to have all the benefits of the questioning account, and then some. The two can share an account of how appraisal processes issue in curiosity, and they can (and should) agree that curiosity automatically activates innate and previously rewarded motor plans. They should also agree that motor plans are evaluated for their costs and their likely reward values. These can then figure in decision-making among options, combining with estimates of likelihood to issue in representations of expected value. But the questioning account postulates, in addition, a representation whose content is a question. This will be at least minimally conceptual in nature, implicating concepts like WHAT, WHERE, WHEN, and WHO. The motoric theory, in contrast, can be wholly nonconceptual. The cues that are appraised to issue in signals of ignorance need not be conceptual (although they can be, of course), and the automatic motor outputs and related sensory forward models can be entirely nonconceptual. So there need be no principled objection to ascribing the representations in question to a wide range of other creatures (even insects).

Moreover, failures of recognition or expectation-formation, as well as model-free signals of ignorance, will be graded and admit of degrees. Then if the corresponding forms of learning also admit of degrees (an infant can become more or less familiar with a new toy; a rat-pup can become more or less familiar with the layout of a maze), the motoric account can readily explain the graded nature of curiosity, and how curiosity about something can gradually fade. As the rat-pup becomes increasingly familiar with its surroundings, the motor-impulse to explore will weaken, as will the anticipated degree of positive valence (the anticipated reward-value) attaching to the sensory forward models produced by those impulses. So the motoric account can smoothly explain the graded nature of curiosity without having to postulate additional representational complexity in the way that the questioning account seemingly needs to.

Someone might claim that even if the motoric account is correct, there will still be a question *implicit* in curiosity, however. One might say that the signals of ignorance together with the motor-tendencies that issue in knowledge-acquisition and reward implicitly encode a question to which the subsequent learning constitutes an answer. But there need be no symbolic structures involved other than those implicated in the appraisals of the input, the resulting signals of ignorance, and the motor-instructions themselves, together with those that result in subsequent learning.

Indeed, one might even say that differing question-forms are implicit in different kinds of curiosity. Thus ignorance of where one is, of the sort that causes widening circles of exploratory behavior, might be said to implicitly pose the question, *what is around here*; failure to recognize a novel object that causes an animal to approach closer, sniff at it, and so on might be said to implicitly ask, *what that is*; and interest in the outcome of a fight that causes sustained attention to the unfolding events might be said to implicitly encode the question, *who will win*. But again, no such questions need to be represented explicitly in symbolic structures of any sort for curiosity to do its work.

More importantly, appealing to implicit questions fails to provide any real explanatory benefit. There is a contrast here with the use made of implicit mental states in other domains, such as generalizations within the mentalizing system (*seeing leads to knowing*), or within the visual system (*light shines from above*). In each of the latter two cases the information in question can be thought of as having been built into a processing-rule rather than carried by a symbolic structure. For instance, the information that seeing leads to knowing can be implicit in the rule that enables a direct transition from JOHN SEES WHAT IS IN THE BOX to JOHN KNOWS WHAT IS IN THE BOX without consulting an explicit major premise with the content, *seeing leads to knowing*. In connection with exploratory curiosity, in contrast, the entire explanation comprises anticipations of exploratory behavior being appraised as attractive together with the actual rewarding qualities of subsequent learning, which in turn reduces the urge to explore. Likewise, mutatis mutandis, for other forms of curiosity. The suggestion that spatial curiosity implicitly poses a question ("What is around here?") is a mere gloss on what really does the explanatory work.

Note that all of the resources appealed to by the motoric theory are either shared by the questioning account or are well-established findings from cognitive science cited earlier. These include automatic motor-activation resulting from affective appraisals, sensory forward models produced from those motor representations, affective evaluation of both of those kinds of representation, and the use of valence as a common currency in decision-making. So the only remaining question is whether the motoric account can "scale up" to incorporate familiar human forms of curiosity (often called "epistemic curiosity;" Kang et al. 2009; Murayama et al. 2019).

4 Scaling Up

We can begin at the beginning, in human infancy. The most basic kind of curiosity-driven behavior is directed attention. Infants in their first few months of life will preferentially attend to stimuli that drive learning. And it seems that the appraisals that issue in their curiosity are, in part, domain-general, responding to any form of novelty, while preferentially attending to novel stimuli of intermediate complexity that issue in sustained learning (Kidd et al. 2012; Poli et al. 2020). But infant curiosity is also partly domain-specific, especially when driven by violations of the expectations created by their core knowledge of objects, agents, set sizes, and more (Spelke 2022). Indeed, the expectancy-violation looking-time paradigm has been extensively used in the first year of life to explore what infants already know and expect about the world (Margoni et al. 2024). Sustained looking after a violation of expectation is a simple form of curiosity.

Infant curiosity is also domain-specific in another sense. It is differentially sensitive to social stimuli of all sorts, especially the infant's native language, faces and face-like stimuli in general, and agency (Johnson et al. 2008; Gervain & Mehler 2010; Spelke 2022). Infants will avidly observe anything they appraise to be an agent, for example (cued by factors such as self-initiated moment). They thereby learn about the individual agents in question, as well as gradually elaborating and extending their core mentalizing system (Baillargeon et al. 2016). The motoric theory can then propose that the underlying mechanism may be a model-free signal of ignorance that outcompetes neural populations representing various possibilities for what the agent will do next. Since agents belong to an especially valued domain, this drives attentional behavior and attaches significant anticipatory reward-value to the forward models of that behavior. The resulting learning is experienced as rewarding, motivating further attention – albeit to a lesser degree when the model-free signal of ignorance is progressively reduced.

Once infants gain more control of their movements and then become toddlers their range of curiosity-driven investigative behavior becomes greatly enriched. They grasp, handle, and manipulate objects, and put them in their mouths; and they can toddle over to look at and play with things that spark their curiosity. Some of these actions may form part of the innate repertoire of the curiosity system, requiring only maturational processes to emerge; others may be a product of familiar forms of reward-based learning. Actions that issue in rewards (in this case caused by a reduction in the model-free signal of ignorance) are repeated, and will thereafter be appraised as rewarding in anticipation.

Now consider how children might learn to ask verbal questions to satisfy their curiosity. On a questioning account, they learn how to directly express both the content and attitude of their questioning state in much the same way as they might learn to express their belief that P by saying, "I believe P" or "I think P." According to the motoric theory, in contrast, children need to acquire a novel set of motor routines – ones that are apt to produce learning (thus reducing or removing the ignorance whose appraisal causes curiosity), thereby issuing in reward-signals that strengthen those motor tendencies themselves. How does this happen?

There are at least two mutually-consistent possibilities. One is that children hear questions being asked that express their own existing curiosity / appraisal of ignorance followed by experiences that satisfy their curiosity. Normal reward-based instrumental learning can then lead them to adopt the same behavior when curious (asking a verbal question). Think of a child being read a story by a care-giver, for example. The protagonist in the story reaches an obvious choice-point. The care-giver says, “What will little-bear do? Let’s turn the page and find out.” Turning the page issues in a resolution, satisfying the child’s own curiosity, and thereby teaching that the “what will …?” question-form is predictive of learning-based reward.[6]

Alternatively (or in addition), observational learning can lead to the same outcome (Dubey et al. 2021). We just have to suppose that young children can identify goals and goal-satisfaction in others (for which there is a good deal of evidence, even in infancy; Baillargeon et al. 2016). For instance, the child observes one care-giver answering a skype-call from another, asking, “When will you be home?” to which the answer is, “Soon.” The first care-giver then replies, “Oh good, so I can start preparing dinner.” Observational learning of the sort we know many animals can use (even bees; Bridges et al. 2023) can lead the child to realize the value of question-asking when ignorant (at least when combined with at least some existing linguistic understanding and mentalizing capacities).

Can the motoric theory scale up even further to explain curiosity about the drivers of global warming, or about the origins of the universe? Notice, first, that it is widely agreed in the psychology literature that curiosity shares the same basic reward-circuitry as do desires and other affective states (Litman 2005; Kang et al. 2009; Blanchard et al. 2015; Kidd & Hayden 2015; Murayama et al. 2019; Sharot & Sunstein 2020). Curiosity both issues in anticipations of reward and then experiences of reward when satisfied. As a result, normal processes of secondary reward-learning will lead to rapid proliferation in the topics that someone is apt to feel curious about, as well as in the types of action that can secure those rewards. (Recall from Section 3 that affect-caused motor urges can initially be quite abstract, like “get closer to that thing.”) And these will vary across individuals with accidents of experience (which topics are discussed at

---

[6] Children will also hear people expressing their curiosity in metacognitive terms, of course (arguably reflecting our common-sense conception of curiosity), saying things like, “I want to know what little-bear does next, let’s turn the page and find out.” So they can learn to express curiosity in that way, too.

home, which topics are learned successfully in school, and so on). Moreover, topics one learns about will tend issue in a sort of snowball effect, resulting in yet more curiosity (Ten et al. 2024). Since the initial learning is to some degree rewarding, that will strengthen the anticipatory reward attached to actions that might issue in further learning about the topic, and so on, until there is not much more to learn.

The motoric theory can thus explain how topics that evoke curiosity (especially sustained curiosity) vary across people with accidents of individual history, as well as perhaps with variations in the initial innate settings of the various drivers of infant curiosity. But can it explain how curiosity can be sparked spontaneously and endogenously, in addition to being provoked by external stimuli of various sorts? This is harder, but arguably can draw on the same basic framework used to explain mind-wandering, in which ideas and thoughts can also be spontaneous and endogenously caused (Author 2025).

Mind-wandering seems to depend on the ventral attentional network, which links regions around the temporoparietal junction with the anterior insula and ventrolateral prefrontal cortex, especially in the right hemisphere (Corbetta et al. 2008; Uddin 2015; Fiebelkorn & Kastner 2020). This evaluates thoughts and memories that have been associatively activated unconsciously by one's ongoing perceptual and cognitive processes, evaluating them for relevance to one's goals and values, and competing for control of the dorsal attentional network. When that competition is won, the representations in question enter working memory and one finds oneself entertaining a new conscious thought. Then among the representations evaluated within the ventral attentional network are likely to be predictions from newly-activated ideas that get out-competed by a model-free ignorance signal, as well as competing abstract motor representations of potential investigative actions. Hence one can find oneself spontaneously in a novel state of curiosity.

Finally, can the motoric account explain how curiosity can motivate people to experience things they know to be aversive (even highly aversive, like the gruesome aftermath of a bad car accident)? It can, via two independent (albeit often co-occurring and interacting) routes. One is that highly negative events, because of their adaptive significance, are likely to issue in an especially powerful and urgent ignorance-signal, which in turn drives a powerful motor response. This might happen swiftly enough to evade executive control. But even if not, suppressing the response will be appraised as effortful (Kool & Botvinick 2013). Resisting the urge to look will be unpleasant (just as resisting an urge to scratch can be), and this can outweigh the anticipated negative experience that will result, resulting in a decision to look. In addition (and for the same evolutionary reason), negative events are apt to be anticipated to be highly informative, and so highly rewarding. These anticipated rewards may be sufficient to outweigh the expected negative feelings that will accompany the experience itself (Oosterwijk et al. 2020). Indeed, such mixed valences can help explain why curiosity is sometimes characterized as a positive emotion, sometimes as negative.

5 Conclusion

This paper has sketched a novel – motoric – theory of curiosity, to be evaluated and compared with existing metacognitive and questioning accounts. The proposed theory entails that curious creatures need not be self-aware in any rich – mentalizing-involving – sense, thus permitting states of curiosity to be widespread across the animal kingdom. In this it resembles the questioning account, and (like the latter) contrasts with standard metacognitive theories. Also like the questioning account, it might best be developed in such a way as to include an appeal to "model free" *de re* signals of ignorance in the appraisal processes that issue in curiosity. Yet the motoric theory can drop any need for explicitly / symbolically represented questions to provide content for one's curiosity. Rather, the contents of curiosity are valenced representations of automatically activated motor instructions and their likely consequences. At the same time, since all of the representations involved can be nonconceptual / analog-magnitude ones, the motoric-theory can readily explain how creatures lacking even the simplest concept-like forms of representation can be curious, as well as how curiosity can admit of degrees and fade gradually with learning.

Acknowledgments

I am grateful to person number one, person number two, person number three, person number four, Kou Murayama, the University philosophy-of-mind reading group, and an anonymous referee for their comments and advice on earlier versions of this essay.

References

Adolphs, R. & Anderson, D.J. (2018). *The Neuroscience of Emotion*. Princeton University Press.

Author, A.N. (2025). *Three Word Title: A sub-title of five words.* A book publisher.

Baillargeon, R., Scott, R., & Bian, L. (2016). Psychological reasoning in infancy. *Annual Review of Psychology*, 67, 159–186.

Baranes, A., Oudeyer, P.-Y., & Gottlieb, J. (2014). The effects of task difficulty, novelty and the size of the search space on intrinsically motivated exploration. *Frontiers in Neuroscience*, 8, 317.

Bennett, D., Bode, S., Brydevall, M., Warren, H., & Murawski, C. (2016). Intrinsic valuation of information in decision making under uncertainty. *PLoS Computational Biology*, 12, e1005020.

Blanchard, T., Hayden, B., & Bromberg-Martin, E. (2015). Orbitofrontal cortex uses distinct codes for different choice attributes in decisions motivated by curiosity. *Neuron*, 85, 602–614.

Bridges, A., MaBouDi, H., Procenko, O., Lockwood, C., Mohammed, Y., Kowalewska, A., Romero González, J., Woodgate, J., & Chittka, L. (2023). Bumblebees acquire alternative puzzle-box solutions via social learning. *PLoS Biology*, 21, e300219.

Bromberg-Martin, E. & Hikosaka, O. (2009). Midbrain dopamine neurons signal preference for advance information about upcoming rewards, *Neuron*, 63, 119–126.

Carruthers, P. (2018). Basic questions, *Mind and Language*, 22, 130–147.

Carruthers, P. (2021). The contents and causes of curiosity. *British Journal for the Philosophy of Science*. Online first: doi/10.1086/716609

Carruthers, P. (2024). *Human Motives: Hedonism, Altruism, and the Science of Affect.* Oxford University Press.

Cheeseman, J., Millar, C., Greggers, U., Lehmann, K., Pawley, M., Gallistel, C., Warman, G., & Menzel, R. (2014). Way-finding in displaced clock-shifted bees proves bees use a cognitive map. *Proceedings of the National Academy of Sciences*, 111, 8949–8954.

Chib, V., Rangel, A., Shimojo, S., & O'Doherty, J. (2009). Evidence for a common representation of decision values for dissimilar goods in human ventromedial prefrontal cortex. *The Journal of Neuroscience*, 29, 12315–12320.

Cisek, P. & Kalaska, J. (2010). Neural mechanisms for interacting with a world full of action choices. *Annual Review of Neuroscience*, 33, 269–298.

Corbetta, M., Patel, G., & Shulman, G. (2008). The reorienting system in the human brain: From environment to theory of mind. *Neuron*, 58, 306–324.

Cos, I., Duque, J., & Cisek, P. (2014). Rapid prediction of biomechanical costs during action decisions. *Journal of Neurophysiology*, 112, 1256–1266.

Daddaoua, N., Lopes, M., & Gottlieb, J. (2016). Intrinsically motivated oculomotor exploration guided by uncertainty reduction and conditioned reinforcement in non-human primates. *Nature: Scientific Reports*, 6, 20202.

Dubey, R., Mehta, H., & Lombrozo, T. (2021). Curiosity is contagious: A social influence intervention to induce curiosity. *Cognitive Science*, 45, e12937.

Dufau, S., Grainger, J., & 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, 1117–1128.

Dunlosky, J. & Metcalfe, J. (2009). *Metacognition*. Sage.

Evans, G. (1982). *Varieties of Reference*. Oxford University Press.

Fiebelkorn, I. & Kastner, S. (2020). Functional specialization in the attention network. *Annual Review of Psychology*, 71, 221–249.

Foley, R. (1987). *The Theory of Epistemic Rationality*. Harvard University Press.

Forstmann, B., Ratcliff, R., & Wagenmakers, E-J. (2016). Sequential sampling models in cognitive neuroscience: Advantages, applications, and extensions', *Annual Review of Psychology*, 67, 641–666.

Fortes, I., Vasconcelos, M., & Machado, A. (2016) Testing the boundaries of "paradoxical" predictions: Pigeons do disregard bad news, *Journal of Experimental Psychology: Animal Learning and Cognition*, 42, 336–346.

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

Frijda, N. (2010). Impulsive action and motivation. *Biological Psychiatry*, 84, 570–579.

Gervain, J. & Mehler, J. (2010). Speech perception and language acquisition in the first year of life. *Annual Review of Psychology*, 61, 191–218.

Gibbons, M., Versace, E., Crump, A., Baran, B., & Chittka, L. (2022). Motivational trade-offs and modulation of nociception in bumblebees. *Proceedings of the National Academy of Sciences*, 119, e2205821119.

Gilbert, D. & Wilson, T. (2007). Prospection: Experiencing the future. *Science*, 317, 1351–1354.

Gilbert, D. & Wilson, T. (2009). Why the brain talks to itself: Sources of error in emotional prediction. *Philosophical Transactions of the Royal Society B*, 364, 1335–1341.

Gipson, C., Alessandri, J., Miller, H.C., & Zentall, T. (2009). Preference for 50% reinforcement over 75% reinforcement by pigeons, *Leaning and Behavior*, 37, 289–298.

Goldman, A. (1999). *Knowledge in a Social World*. Oxford University Press.

Gruber, M., Gelman, B., & Ranganath, C. (2014). States of curiosity modulate hippocampus-dependent learning via the dopaminergic circuit, *Neuron*, 84, 486–496.

Goupil, L. & Proust, J. (2023). Curiosity as a metacognitive feeling. *Cognition*, 231, 105325.

Heyes, C. (2015). Animal mindreading: What's the problem? *Psychonomic Bulletin and Review*, 22, 313–327.

Jeannerod, M. (2006). *Motor Cognition*. Oxford University Press.

Johnson, M.H., Grossmann, T., & Farroni, T. (2008). The social cognitive neuroscience of infancy: Illuminating the early development of social brain functions. *Advances in Child Development and Behavior*, 36, 331–372.

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

Karttunen, L. (1977). Syntax and semantics of questions. *Linguistics and Philosophy*, 1, 3–44.

Khalvati, K., Kiani, R., & Rao, R. (2021). Bayesian inference with incomplete knowledge explains perceptual confidence and its deviations from accuracy. *Nature communications*, 12, 5704.

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

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

Krieglmeyer, R., Deutsch, R., De Houwer, J., & De Raedt, R. (2010). Being moved: Valence activated approach-avoidance behavior independently of evaluation and approach-avoidance intentions. *Psychological Science*, 21, 607–613.

LeDoux, J. (2012). Rethinking the emotional brain, *Neuron*, 73, 653–676.

Kool, W. & Botvinick, M. (2013). The intrinsic cost of cognitive control. *Behavioral and Brain Sciences*, 36, 697–698.

Le Pelley, M. (2012). Metacognitive monkeys or associative animals? Simple reinforcement learning explains uncertainty in nonhuman animals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38, 686–708.

Levy, D. & Glimcher, P. (2012). The root of all value: A neural common currency for choice. *Current Opinion in Neurobiology*, 22, 1027–1038.

Litman, J. (2005). Curiosity and the pleasures of learning: Wanting and liking new information, *Cognition and Emotion*, 19, 793–814.

Loewenstein, G. (1994). The psychology of curiosity: A review and reinterpretation. *Psychological Bulletin*, 116, 75–98.

Margoni, F., Surian, L., & Baillargeon, R. (2024). The violation-of-expectation paradigm: A conceptual overview. *Psychological Review*, 131, 716–748.

McDowell, J. (1994). *Mind and World*. Harvard University Press.

McNamara, J. & Houston, A. (1986). The common currency for behavioral decisions. *The American Naturalist*, 127, 358–378.

McNamee, D. & Wolpert, D. (2019). Internal models in biological control. *Annual Review of Control, Robotics, and Autonomous Systems*, 2, 339–364.

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., & Watzl, S. (2005). Honey bees navigate according to a map-like spatial memory. *Proceedings of the National Academy of Sciences*, 102, 3040–3045.

Mischiati, M., Lin, H-T., Herold, P., Imler, E., Oldberg, R., & Leonardo, A. (2014). Internal models direct dragonfly interception steering. *Nature*, 517, 333–338.

Mishra, S. (2014). Decision-making under risk: Integrating perspectives from biology, economics, and psychology. *Personality and Social Psychology Review*, 18, 280–307.

Montague, P.R. & Berns, G. (2002). Neural economics and the biological substrates of valuation. *Neuron*, 36, 265–284.

Murayama, K., FitzGibbon, L., & Sakaki, M. (2019). Process account of curiosity and interest: A reward-learning perspective. *Educational Psychology Review*, 31, 875–895.

Nicholson, T., Williams, D.M., Grainger, C., Lind, S., & Carruthers, P. (2019). Relationships between implicit and explicit uncertainty monitoring and mindreading: Evidence from autism spectrum disorder. *Consciousness and Cognition*, 70, 11–24.

Nicholson, T., Williams, D.M., Grainger, C., Lind, S., & Carruthers, P. (2021). Linking metacognition and mindreading: Evidence from autism and dual-task investigations. *Journal of Experimental Psychology: General*, 150, 206–220.

Oosterwijk, S., Snoek, L., Tekoppele, J., Engelbert, L., & Scholte, H.S. (2020). Choosing to view morbid information involves the reward circuitry. *Nature Scientific Reports*, 10, 15291.

Penn, D., Holyoak, K., & Povinelli, D. (2008). Darwin's mistake: Explaining the discontinuity between human and nonhuman minds. *Behavioral and Brain Sciences*, 30, 109–178.

Pleskac, T. & Busemeyer, J. (2010). Two-stage dynamic signal detection: A theory of choice, decision time, and confidence, *Psychological Review*, 117, 864–901.

Poli, F., Meyer, M., Mars, R., & Hunnius, S. (2022). Contributions of expected learning progress and perceptual novelty to curiosity-driven exploration. *Cognition*, 225, 105119.

Poli, F., Serino, G., Mars, R., & Hunnius, S. (2020). Infants tailor their attention to maximize learning. *Science Advances*, 6, eabb5053.

Ruff, C. & Fehr, E. (2014). The neurobiology of rewards and values in social decision-making. *Nature Reviews Neuroscience*, 15, 549–562.

Rupert, R. (2018). Representation and mental representation. *Philosophical Explorations*, 21, 204–225.

Seeley, T. (1995). *The Wisdom of the Hive*. Harvard University Press.

Seligman, M., Railton, P., Baumeister, R., & Sripada, C. (2016). *Homo Prospectus*. Oxford University Press.

Sescousse, G., Li, Y., & Dreher, J-C. (2015). A common currency for the computation of motivational values in the human stratum. *Social Cognitive and Affective Neuroscience*, 10, 467–473.

Sharot, T. & Sunstein, C. (2020). How people decide what they want to know, *Nature Human Behavior*, 4, 14–19.

Shea, N. (2018). *Representation in Cognitive Science*. Oxford University Press.

Shizgal, P. & Conover, K. (1996). On the neural computation of utility. *Current Directions in Psychological Science*, 5, 37–43.

Spelke, E. (2022). *What Babies Know*. Oxford University Press.

Ten, A., Oudeyer, P., Sakaki, M., & Murayama, K. (2024). The curious U: Integrating theories linking knowledge and information-seeking behavior. *Manuscript under review*.

Teodorescu, A. & Usher, M. (2013). Disentangling decision models: From independence to competition, *Psychological Review*, 120, 1–38.

Uddin, L. (2015). Salience processing and insular cortical function and dysfunction. *Nature Reviews Neuroscience*, 16, 55–61.

Usher, M. & McClelland, J. (2001). The time course of perceptual choice: The leaky, competing accumulator model, *Psychological Review*, 108, 550–592.

van den Berg, R., Anandalingam, K., Zylberberg, A., Kiani, R., Shadlen, M., & Wolpert, D. (2016a). A common mechanism underlies changes of mind about decisions and confidence. *eLife*, 5, e12192.

van den Berg, R., Zylberberg, A., Kiani, R., Shadlen, M., & Wolpert, D. (2016b). Confidence is the bridge between multi-stage decisions. *Current Biology*, 26, 3157-3168.

Vasconcelos, M., Monteiro, T., & Kacelnik, A. (2015). Irrational choice and the value of information. *Nature: Scientific Reports*, 5, 13874.

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

Whitcomb, D. (2010). Curiosity was framed, *Philosophy and Phenomenological Research*, 81, 664–687.

Williamson, T. (2000). *Knowledge and its Limits*. 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, 1573–1576.

Wolpert, D. & Ghahramani, Z. (2000). Computational principles of movement neuroscience. *Nature Neuroscience*, 3, 1212–1217.

Wolpert, D. & Kawato, M. (1998). Multiple paired forward and inverse models for motor control. *Neural Networks*, 11, 1317–1329.