Rev.Phil.Psych.
https://doi.org/10.1007/s13164-017-0372-1

The Epistemology of Rational Constructivism

Mark Fedyk[1] · **Fei Xu**[2]

Abstract Rational constructivism is one of the leading theories in developmental psychology. But it is not a purely psychological theory: rational constructivism also makes a number of substantial epistemological claims about both the nature of human rationality and several normative principles that fall squarely into the ambit of epistemology. The aim of this paper is to clarify and defend both theses and several other epistemological claims, as they represent the essential epistemological dimensions of rational constructivism.

1 Introduction

Rational constructivism is a leading theory in developmental psychology. It aims to describe the psychological mechanisms that guide the formation of children's early knowledge of their physical and social worlds. And like many of the best theories in psychology, most of the evidence for rational constructivism is experimental (Xu 2007, 2016; Xu and Kushnir 2012, 2013; Xu and Tenenbaum 2007; Gopnik and Wellman 2012). But as the name suggests, rational constructivism is neither a purely descriptive nor a purely empirical theory: a number of normative claims about reasoning and knowledge acquisition are essential to rational constructivism. That is true, partly, because rational constructivism is a theory of what forms of rationality characterize the human mind, and also a theory of how rationality might be implemented in the mind-brain system. But, because of that, rational constructivism is also a theory—albeit only a partial, defeasible, and at most approximately true theory—of how an agent

✉ Mark Fedyk
mfedyk@mta.ca

✉ Fei Xu
fei_xu@berkeley.edu

[1] Department of Philosophy, Mount Allison University, Sackville, Canada

[2] Department of Psychology, University of California, Berkeley, USA

Published online: 27 November 2017

should reason in contexts where the acquisition of knowledge is possible. And this means that rational constructivism is as much an epistemological theory as it is a psychological theory..[1]

An immediate clarification is called for. The question which motivates our philosophical interest in rational constructivism is: What concepts and beliefs, capabilities and principles, do or should people have, in virtue of which they are able to learn? By saying that rational constructivism is (partly) an epistemological theory, our specific claim is that rational constructivism provides a limited, but nevertheless substantive, answer to this question. What's more, we note that an intuitively plausible analytic definition of knowledge will not be a psychologically sufficient answer to this question. There must be a much richer set of representations and abilities that learners have in virtue of which they are able to confront the world and come away, often enough, with knowledge—and psychology must play a central role in the project of uncovering the relevant details (Quine 1969).

The psychological dimensions of rational constructivism are well known by now (Xu and Griffiths 2011; Xu and Kushnir 2012). However, little work has been done to clarify and defend the epistemological dimensions of rational constructivism; doing just that is the aim of this paper. And the epistemology of rational constructivism can be divided into what might be called the "epistemological core" and "epistemological periphery" of rational constructivism. The core consists of the (again, only partial) theory of rationality that is essential to rational constructivism. The periphery refers to a cluster of epistemological precepts that are, on instrumental grounds, linked with the core of rational constructivism. Put another way, the epistemological periphery of rational constructivism consists of a number of practical norms that would, if acted upon, facilitate knowledge acquisition *if* much of the psychological *and* epistemological core of rational constructivism is largely correct.

This paper is organized as follows. The psychological core of rational constructivism is described in some detail in section 2.0. We then turn to the epistemological core in section 3.0, and there our aim is to show how the concept *cognitive agency* can be derived from rational constructivism's distinctive conception of rationality. We then shift our attention, in section 4.0, to the epistemological periphery of rational constructivism. In this section we show how the aforementioned epistemological norms are linked with cognitive agency. A handful of concluding philosophical observations are offered in the final section 5.0.

2 Rational Constructivism

Rational constructivism was developed in the early 2000s as an alternative to its two primary competitors in psychology, the families of nativist and empiricist theories of cognition that have been developed by the cognitive sciences [cf. (Haith 1998; Spelke 1998)]. It also aims to develop a constructivist account of cognitive development that is non-Piagetian (Piaget 1954). In brief, rational constructivism rejects the nativist view

[1] We are exceptionally grateful to Tyler Burge, Roberto Casati, David Danks, Tania Lombrozo, Tamar Kushnir, and one anonymous reviewer at this journal for the helpful comments, suggestions, criticisms, and philosophical insights they provided to us during the process of writing this article.

that many of the most common and most reliable concepts and inferential patterns that people use in their daily lives are fixed by cognitive information and processes that are almost entirely developmentally endogenous. But rational constructivism also rejects the empiricist view that all of the very same everyday mental representations are acquired by copying them from the perceptual representations that form the content of sensory experience [cf. (Fodor 2005)]. What makes rational constructivism a truly novel alternative is that, as we will soon explain, according to rational constructivism, a crucial number of the mental representations used by cognition are neither empirical generalizations from experience nor given innately. Rational constructivism posits another source for at least some of a learner's most useful concepts and inferential tendencies: these mental representations are *constructed* by the learner herself [cf. (Carey 2009; Piaget 1954; Quine 1960)].

To prevent misunderstanding, it is important for us to emphasize that rational constructivism has a purely *computational* ontology (Marr 2010). The theory of rational constructivism should be interpreted as making claims only about the structure and the content of the flow of information that *is* human cognition, and not also about the sensory and neurophysiological mechanisms that implement the relevant flow. Our talk of mechanisms, then, should be taken as referring, restrictively, to computational mechanisms [cf. (Otto and Rusanen 2011; Griffiths et al. 2010; Tenenbaum et al. 2011)]. Important work remains to be done determining how the computational processes are implemented in psychological processes [cf. (Jones and Love 2011)].[2]

Here now are the main scientific theses of rational constructivism [cf. (Xu 2016)]:

1. **The Starting State:** Humans are born with a suite of *proto-conceptual* primitives. These early informational structures may not have all of the traditional properties usually attributed to fully lexicalizable concepts, but they go beyond sensory representations. Infants begin with enough knowledge to constrain and organize initial cognition, and the evidence to date indicates that these early information structures do not have all the properties of a Fodorian language of thought, although unlike Fodorian modules, they are independent from specific sensory modalities. Furthermore, it is also likely that these early informational structures undergo substantial developmental change during even the earliest stages of cognitive development.
2. **Lexicalization and Concepts:** A categorical change occurs early in cognitive development according to which concepts acquire symbolic and lexical structure. Lexicalization can subsequently lead to radical and systematic belief revision, as well as the capacity to form a much wider array of beliefs. Importantly, these new concepts are neither empirical generalizations of sensory experience nor statistical elaborations of endogenous proto-conceptual structures. To a good first approximation, they express (in structured and symbolic form) guesses (that become progressively more educated) about such things as object, number, causality, and agency. In the first instance, these concepts are constructs formed by cognition that are then conserved, refined, or abandoned as a result of conflicts with information learned

[2] But also, as our arguments will show in due course, we reject such overly strong assumptions that *all* cognition is intrinsically rational according to *only* Bayesian standards. In this respect we are innocent of the charge of being Bayesian fundamentalists, as per (Jones and Love 2011).

from experience. But during and after lexicalization, the initial concepts continue to function within the cognitive system—the non-symbolic concepts of object, number and causality, for example, provide scaffolds that support learning in the corresponding domains [cf. (Carey 1987, 2011; Spelke 2003; Xu 1997, 1999)].

3. **Multiple Learning Mechanisms:** There are at least three different inferential processes that support learning: language learning, Bayesian induction, and constructive thinking. Language learning provides cognition with symbolic structures and templates useful for encoding information, and it also constraints the representational repertoire of any one mind so that this repertoire is a reasonably good match of the repertoire of other minds—namely, the minds of other speakers of the same or similar dialect(s). Language learning thus provides a library of labels (a way of symbolically representing) the objects of (possible) experience. Bayesian learning provides a mechanism for increasingly reliable inductive inference based on data provided by experience. Bayesian learning is hypothesized to play an important role in belief fixation—such that representations with a sufficiently high posterior probability become a person's beliefs. Future evidence may of course lower the probability of these beliefs, leading to, potentially quite radical change in belief (see below). However, neither Bayesian reasoning nor imitative language use are particularly rich sources of novel hypotheses that can be tested against experience. Constructive thinking fills this gap, as it refers to the ability of the mind to construct novel explanations, imagine alternative mechanisms, make analogies, engage in thought experiments, and, generally speaking, enlarge a mind's conceptual repertoire by constructing new ideas ("learning by thinking", Lombrozo 2012; Gendler 2000; Gentner and Hoyos 2017). Constructive learning thus amplifies the efficacy of the other two learning mechanisms.
4. **Radical Conceptual Change Is Possible:** Radical conceptual change is not only always possible according to rational constructivism; it sometimes can be necessary in order to increase the fidelity of a learner's individual store of knowledge. The first and most obvious instance of this is the early lexicalization of proto-conceptual primitives that occurs as young children learn their first language. But more generally, the learning mechanisms posited by rational constructivism do not prevent—and under the right conditions can facilitate—radical conceptual change. At no point is any of a learner's beliefs—no matter how central it is—immune to revision or even rejection.
5. **Forming Beliefs Involves Probabilistic Hypothesis Testing:** According to rational constructivism, to acquire knowledge is to synthesize empirical information with constructed ideas by a process of approximately Bayesian belief fixation and revision. For that reason, learning can be characterized, albeit only at the level of idealization, as a recurring sequence that moves from some prior probability distribution over a set of hypotheses, the computation of the posterior probabilities of these hypotheses given the strength of evidence as dictated by Bayes' Theorem, and then the acceptance of those beliefs determined to be most probable at that time. The most probable ideas in the mind of a learner at some time—whether these ideas are acquired from testimony, observation, or constructive thinking—are the learner's beliefs.
6. **Learning is an Active, Agentive, and Social Process:** Starting in infancy, many of the ideas that are subject to rational evaluation by the probabilistic hypothesis

testing mechanism are at least partially by-products of a child's own process of data generation [cf. (Harris 2000; Frazier et al. 2009; Cook et al. 2011; Sim and Xu 2017; Walker et al. 2015; Weisberg 2016)], which is of course the *constructive* element of rational constructivism. Imagination, individual and social play, pretend world-making, increased language learning, question asking, seeking help, and various other social, linguistic, and cognitive phenomena are important sources of new information that all support learning.

It is also important to emphasize *how* language learning and Bayesian induction work according rational constructivism's theory. In the case of language, it provides not only symbols but also *formal properties* that have to be satisfied in order for learners to be able to use language coherently. For instance, certain rules of syntax take the form of hypothetical imperatives for language learners: in order to understand and be understood, these particular formal bits of language (noun phrases, generic terms) *must* be used in such-and-such a way. But this view of language learning must be tempered with the parallel view of how *agentive* language learning can be. Word learning can also supply a learner with a set of algorithms that, through their proactive application, radically extend a learner's conceptual knowledge (e.g., Waxman and Markow 1995; Welder and Graham 2001; Xu 2002). And, most importantly, language also provide a very nearly overwhelming amount of information about what *needs to be learned* in order to count as a knower. Because of its basis in language, the standard of what counts as a knower is frequently community-based, and so it is an implication of rational constructivism's theory that an important amount of active social learning interacts with an equally active process of language learning.

Bayesian induction is the learning mechanism that, by rational constructivism's light, is ultimately responsible for fixing belief. Formally, rational constructivism says that the computations which drive belief fixation most likely involve the construction of Bayesian probabilistic models that are subsequently evaluated by an algorithm which applies some formulation of Bayes' rule. The probabilistic, causal, generative models represent information about the external world as directed acyclical graphs (DAG) (Danks 2014). Each DAG is defined by a set of nodes and edges. Nodes are variables that are standardly interpreted as representing properties, while the edges running between the nodes are standardly interpreted as representing causal relations between the relevant properties. By hypothesis, the structure of any discrete DAG can be used to define a joint probability distribution over all of the variables in the model, which in turn can be used to make inferences—for example, it may entail that two variables in the model are perfectly negatively correlated with one another. Each model can then be treated as a unique hypothesis about the causal behaviour realized by a set of properties. Learning is then construed as, at some substantive level of psychological organization, a process essentially involving multiple different hypotheses (i.e., DAGs) that are formed and then evaluated statistically. Whichever hypothesis is found to have the highest posterior probability becomes a belief. So, technically speaking, rational constructivism offers an alternative theory of belief formation from both the connectionist models popular amongst empiricists (according to which beliefs are defined as vectors of activation patterns in connectionist networks) and the traditional symbolic models popular amongst nativists (according to which beliefs are syntactic composites formed out of innate atomistic lexical concepts) (e.g., Griffiths et al. 2010; Tenenbaum et al. 2011).

According to rational constructivism, then, the knowledge that even a very young child acquires is neither a generalization of empirical patterns in her experience (empiricism) nor a copy of lexical representations acquired either from learning her parent's language or as a matter of developmentally-endogenous maturation (nativism). Some of her ideas are truly her own, and some of these ideas may subsequently be accepted as beliefs because they are probable on the basis of her experience. As a child grows and her cognition becomes more sophisticated, rational constructivism claims that her knowledge will be an increasingly complex synthesis of her physical experience, the by-products of her own constructive processes of thought, and the information she receives from various different kinds of social interaction.

Someone may object at this point that even this story is too simple. For example, in addition to supplying the mind with formal templates that are useful for constructing generalizable knowledge, language also supplies learners with a set of hypothetical imperatives that tell them, inter alia, how they should think about certain topics. Nevertheless, in order to keep things focused, we have very little more to say about language and the important role it plays in facilitating learning. Likewise, we will have nothing to say about any of the other learning mechanisms, like perception, which also support learning. We do this only in order to ensure that this paper remains appropriately focused, and not because of any judgement on our part about the comparative psychological or epistemological importance of these other mechanisms.

3 The Epistemological Core: The Complex Meaning of "Rational" in Rational Constructivism

Rational constructivism's view that a Bayesian mechanism ultimately fixes beliefs ensures that, at least over the long run, people will tend to believe only those propositions that are most probable given the evidence that they have received up to that point in time. However, it would be a mistake to conclude that this kind rationality is the entirety of what *rational* means within rational constructivism.

One reason for this is that rational constructivism holds that there are a number of sub-personal learning mechanisms, including those which implement the relevant Bayesian inferences. So, we need a conception of rationality that can apply to both personal and sub-personal psychological processes. But another reason is that, as we shall argue, the rationality of constructive thinking is fundamentally different than the rationality of Bayesian inference. We need, therefore, a more expansive conception of rationality. And the conception of rationality which be used to say that x is rational to the extent that it approaches some ideal limit y meets these two criteria. This conception can be applied at different levels of psychological organization, and by substituting in different ideals, it can be used to generate different specifications of rationality.

However, the differing rationalities of the mind's learning mechanisms are not the only complication involved in working out the meaning of *rational* for rational constructivism. Another comes from the fact that, at the personal and conscious level of psychological organization, information from both Bayesian and constructive thinking, as well as other cognitive mechanisms, flows together and thereafter can form the basis of second-order intentions, plans, goals, self-knowledge, and so on. And since it is also possible at this second-order level of psychological organization to deliberate about both

the information generated by the relevant sub-personal mechanisms and whatever other mental states are consciously accessible, there is another conception of rationality that can be located within rational constructivism—one that, because of its personal, conscious, and intentional (in short: *agentive*) dimensions, is closer to what philosophers have traditionally regarded as *the* concept of rationality [cf. (Davidson 2001)].

One further clarification is called for. It is true that philosophers typically apply the concept *rational* to agentive psychological process, but it is nevertheless the case that in so doing philosophers have not always had the same definition of rationality in mind. Repurposing a distinction that Michael Rescorla (Rescorla 2013, p. 472) uses to speak of the normativity of logic, we can distinguish between the following senses of rationality:

- Rationality is normative: in the sense of defining a set of regulative ideals. Certain logical or statistical or inferential principles dictate how one should think, and someone is rational to the extent that their thinking conforms to these ideals.
- Rationality is descriptive: in the sense of a set of regulative ideals that describes how people actually do think, at least to a rough first approximation.
- Rationality is constitutive of thought: conformity to some of the regulative ideals of rationality is, at least in part, just what it is to think.

Note that, as these definitions are formulated, they can only be applied to agents. However, the first two can be easily reformulated so that they apply to subpersonal psychological processes as well. For example, in the first definition the appropriate change can be made by replacing "how one should think" with "how some psychological mechanism should process information". In what follows, we will implicitly be making the relevant reformulations so that the definitions can be applied to the subpersonal psychological processes that are characteristic of Bayesian learning and constructive thinking respectively; this is consistent with standard practice in psychology (Stanovich 2012; Stanovich and West 1998).

That said, we are not also putting forward the proposition that, in substituting personal concepts for process concepts, we are performing a perfectly neutral reformulation. As Zoe Drayson has shown (Drayson 2012), it is frequently better to think of the distinction between personal and subpersonal psychological states as a distinction between doxastic and non-doxastic states (Drayson here follows (Stich 1978)). However, in our case, we are most interested in those psychological states—whether they be personal or not, doxastic or not—that can be assessed for their rationality in any of the three senses specified above. It is another step to conclude that these assessments generate easy conclusions about how people (or their subpersonal processes) are to be held responsible for these states. This is a concern that is frequently in the background in discussions of rationality, and it is also a concern that we address below. But to foreshadow the issue slightly, we do not hold that people are directly responsible for the rationality of their subpersonal (or non-doxastic) psychological processes.

A further clarification may be useful at this juncture. José Bermúdez uses the label "the interface problem" for the work we need to do to clarify how explanations in commonsense psychology and explanations in cognitive science are related (Bermudez 2005). Bermúdez defines four different types solutions to the interface problem: the

autonomous mind, which holds that explanations at the personal and subpersonal level will be categorically different; the *functional mind*, which holds that there is no categorical distinction between personal and subpersonal mental processes, and that all mental processes, which themselves just are mental states like beliefs, desires, and other such mental kinds, are all functional states of the brain; the *representational mind*, which holds that thinking just is a complex set of formal operations defined over a physically implemented "language of thought"; and, finally, the *neurocomputational mind*, which holds that belief-desire psychology should be abandoned and replaced with neuroscientific explanations, as these explanations hold the promise of better accounts of human behavior than any psychological theory which posits beliefs, desire, and other intentional states. How does our analysis of rationality fall into this taxonomy? Since we are restricting our focus to only those psychological states and processes which can be realized as computations, our theory is ontologically incompatible with the neurocomputational mind response to the interface problem. However, there are no further substantial ontological claims built into either rational constructivism or our analysis of its epistemology that render either incompatible with the remaining three types of solution to the interface problem. However, we are independently skeptical of solutions falling within the basket of *representational mind* solution on empirical grounds. Finally, while we believe that there is a real difference between the personal (or doxastic) and subpersonal level of psychological organization, we (as we will soon demonstrate) reject the additional methodological view that concepts which apply to one level of organization cannot be applied to the other. But, because of this, our view most likely should not be categorized as an instance of an *autonomous mind* solution to the interface problem. All that remains, then, is the conclusion that our theory is best treated as a non-standard version of the *functional mind* response to the interface problem.

So, with both the distinction between the three conceptions of rationality and some associated philosophical clarifications in hand, we can now proceed with the task of sorting out what *rational* means for rational constructivism. To this end, we need to distinguish between three different, though interacting, forms of rationality. Each characterizes what it is for certain psychological mechanisms or capacities to be functioning *appropriately* or *correctly*—and so each sense of rationality marks out a liminal space in human cognitive ontology where normative and causal forces interact and even can, some of the time, run parallel with one another.

3.1 Bayesian Rationality

Let us say that the psychological mechanisms that rely upon some kind of Bayesian algorithm to determine which propositions to accept as beliefs exhibit *Bayesian rationality,* to the extent that these mechanisms function *appropriately*—which here means that they typically fix as beliefs only those propositions which have the highest posterior probability. Bayesian rationality is therefore sub-personal, normative, and descriptive. The Bayesian belief-fixing mechanism posited by rational constructivism is not itself under executive or agentive control, and the axioms of probability plus Bayes' theorem are most likely an approximation of whatever algorithm is actually implemented by the relevant mechanisms, and of course the axioms of probability and Bayes's theorem are a set of a regulative ideals.

At this point, a subtle complication arises. We have distinguished between *normative* and *descriptive* sense of rationality, and as this distinction was written out above, we used the phrase "a set of regulative ideals" in definition of both concepts. The complication is that it need not be the same set of regulative ideals that characterizes both normative and descriptive Bayesian rationality. Put more concretely, in the case of the descriptive rationality of Bayesian rationality, the relevant regulative ideals are the axioms of probability plus some formalization of Bayes's rule. But the set of regulative ideals that characterize the normative rationality of the relevant psychological mechanisms is, we suggest, a set of epistemic principles such as: only propositions with a posterior probability greater than some value γ should become beliefs, and if propositions A and B have probabilities greater than γ but it is not possible for both A and B to be true, then whichever proposition is more probable should be believed, and the value of γ should be adjusted upward as the density of a person's web of belief increases. We need this distinction because the axioms of any probability calculus are not themselves meaningful epistemological principles suitable for *explaining* why it is that the human mind is able to acquire knowledge. The axioms are only formal rules, and formal rules provide no explanation of why it would be prudent for a learner (or some part of her mind) to follow them if she wanted to increase her knowledge and minimize her ignorance. But the rules become epistemically meaningful when it can be shown that they are a way of *realizing* or *implementing* certain independently plausible epistemic principles. So, the distinction between the descriptive and normative rationality of the relevant Bayesian belief evaluating mechanisms is crucial because it allows us to describe, albeit only at a very abstract level, how this realization might work, and thereby illustrate one (and certainly not the only) way in which rational constructivism is both an substantive epistemological and psychological theory.

What evidence is there that this complicated conception of Bayesian rationality is psychologically real? The evidence consists primarily of two decades worth of experiments that, taken together, demonstrate that both children and adults make inferences that seem to reflect unconscious knowledge of certain basic principles of logic and probability (e.g., Gopnik and Bonawitz 2015; Griffiths et al. 2008; Oaksford and Chater 2009; Xu and Kushnir 2012, 2013). Crucially, however, these experiments also demonstrate that, even in children as young as four, Bayesian rationality cannot explain *all* of the learning-conducive thinking, despite the fact that it accounts for some otherwise quite sophisticated patterns of judgment.

3.2 Creative Rationality

The second conception of rationality that finds a home in rational constructivism is that of *creative rationality*. Creative rationality has agentive dimensions, and it is a form of normative and descriptive rationality; but it is not constitutive in the relevant sense.

So, what set of regulative ideals does creative rationality bend towards? In fact, it seems that there is only a single principle in the relevant set for both the normative and descriptive dimensions of creative rationality: when the mechanisms responsible for constructive thinking are functioning "well" or "appropriately", they should maximize the number of useful novel ideas that are evaluated for belief by maximizing the total number of novel beliefs that are evaluated, whether or not they are useful. Crucially, "novel" has a technical meaning here: an idea is novel if it is statistically and logically

independent of beliefs that have already been accepted by a learner. And this has a very interesting corollary. Creative rationality can therefore be extremely well expressed by patterns of thought that are deeply *irrational* what evaluated using by the standard of normative Bayesian rationality—such as when a learner spends time entertaining contradictory beliefs, or pursuing a deeply counterfactual scenario the details of which conflict with many of her preexisting beliefs. But all this corollary demonstrates is the utility of employing multiple distinct conceptions of rationality.

As before, there is ample experimental evidence of creative rationality. For example, the precocious imaginative abilities of children are well documented (Harris 2000; Lane et al. 2016; Taylor 2013), and there is also evidence that the capacity to imagine alternative explanatory possibilities plays a deep role in various different forms of learning (Gopnik et al. 2007; Walker et al. 2015). More recent studies also highlights the ability of children to build conceptual structures using self-generated evidence while engaged in self-directed play (Sim and Xu 2017). Finally, there is both good philosophical and compelling scientific evidence that constructive thinking plays an important role in many, many different forms of learning—ranging from the most complicated forms of scientific inquiry to making sense of extremely simple cause and effect relationships (Lombrozo in press).

3.3 Cognitive Agency

The final concept in rational constructivism's theory of rationality is that of *cognitive agency*. It refers to a form of agentive and normative rationality, for which the relevant regulative ideal is a rule saying that one should seek and maintain whatever *balance* between exercising creative rationality and Bayesian rationality maximizes knowledge acquisition. This means that realizing cognitive agency in something close to its ideal form just is to establish a dynamic interplay between expressions of creative rationality and Bayesian rationality that is conducive, over some meaningful period of time, to the formation of knowledge. When it is realized, thus, cognitive agency is a (not total; other factors matter too; so only partial) by-product of a pattern of interaction between creative rationality and Bayesian rationality. This pattern may only occur once, or more likely, may occur repeatedly over long periods of time. The mechanisms that sustain "learning by thinking" and other forms of constructive thought and the mechanisms which implement Bayesian belief evaluation both must be functioning sufficiently *well*—i.e., sufficiently close to the relevant normative ideals—in order for cognitive agency to be realized. And when cognitive agency is realized, it is frequently a source of knowledge.

This idea is, of course, speculative. But it is nevertheless possible to refine it in ways that make it more amenable to empirical investigation. First of all, the relevant meaning that *balance* takes in the definition of cognitive agency is not given by some a priori ratio. We are not suggesting that creative rationality just is to produce ten novel ideas per hour, Bayesian rationality involves accurately determining which of the ten ideas is the most probable and then making that idea into a belief, and that cognitive agency, then, just is maintaining this 10:1 ratio for a sufficiently long period of time. Instead, by definition, the balance that constitutes cognitive agency is achieved whenever other forms of rationality are both realized and interacting in such a way that, often enough, knowledge is produced. Furthermore, the concept of balance we are introducing here

has teleological, causal, and normative components. The balance that *is* cognitive agency is achieved when knowledge acquisition is maximized, and the production of this knowledge is caused by the interaction of different learning mechanisms, and not overly hindered by any exogenous factors (more on this last idea shortly). But again, these learning mechanisms themselves have to be functioning *well*—they themselves must be realizing, albeit imperfectly, their own forms of normative rationality. Each of these dimensions can be the subject of different programmes of empirical research, and this work will shed more light on the ways that cognitive agency can be realized.

Thus, cognitive agency is a form of *cognitive homeostasis*. When the mind's learning mechanisms are functioning appropriately, and interacting with the relevant enduring balance, the effect is to maximize the acquisition of knowledge.

But unlike Bayesian rationality, cognitive agency is close in at least one important respect to what philosophers usually have in mind when they are speaking of rationality: cognitive agency is a form of rationality for which a person can be individually responsible. Because cognitive agency is expressed primarily at the agentive and intentional level of psychological organization, this means that a person's habits of mind, beliefs about their self, their motivational tendencies, and so on can each impact whether the relevant equilibrium is realized. And while things like habits and beliefs cannot themselves be chosen, they nevertheless are influenced, sometimes quite directly and sometimes only incidentally, by deliberation and choice. So, again, cognitive agency is something a learner can be responsible for, in the specific sense that some of their choices can causally influence cognitive agency. And of course, this is not to suggest that one can simply "decide" or "will" themselves into realizing their own cognitive agency; quite probably, it takes a great deal of practice, training, and social support in order to create and then stabilize the relevant equilibrium.

The reason we need to introduce cognitive agency is easy to explain. Neither constructive thinking nor Bayesian inference can, all alone, generates beliefs that are a reasonably accurate picture of the world. Yet the information that both of these learning mechanisms produce is, often enough, necessary in order for a mind to acquire accurate information. Although we do not have room to delve very deeply into the issue, neither language learning nor perception can provide a learner with *all* of the conceptual structures they need to make sense of the world; as we indicated above, at least some of these conceptual structures must be of the learner's own design (e.g, Carey 2009, 2015). But at the same time, a cognitive mechanism that is able to generate novel information will, because of that, rarely be a mechanism that is able to generate *uniformly* accurate information. The output of such a mechanism needs to be filtered. And that is one of the functions that Bayesian rationality plays: it screens for only the most probable ideas, ensuring that an agent's beliefs are only a small subset of all of the ideas that have passed through her mind. But it is also the case that neither of these two learning mechanisms, even when they are operating as close to implementing the relevant regulative ideals as possible, can be individually responsible for producing knowledge, since learning is not a matter of mechanically conjoining the output of all of the mind's different sub-personal mechanisms. Learning is a by-product of the *epistemically-appropriate* integration of the output of constructive thinking and Bayesian inference (and other learning mechanisms too; we are, again, ignoring the other mechanisms simply out of a concern for concision). And the reason why it is crucial that this interaction be regulated or constrained by a further set of epistemic

norms is easy to explain: the information produced by different learning mechanisms is frequently, when aggregated, either inconsistent, incoherent, or incomplete. Learners must therefore do the work of organizing this information into a set of representations that are structured *well-enough*—that is: sufficiently consistent and sufficiently coherent—to count as a reliable representation of the world, or if not that, then decide that inquiry is not yet finished because the available evidence is incomplete and so continuing the search for new evidence or new ideas.

Of course, what norms would confer the appropriate structure on output of the relevant mechanisms is a much more difficult question to answer; we will have a little bit more to say about this particular topic in the next subsection. At this point, it is more helpful to, instead, spend more time examining what cognitive agency looks like when it is realized.

First of all, cognitive agency is something that can have distinctive developmental pathways: it is a capability that can be habituated, fostered, or suppressed by any number of distinct patterns of individual behaviour or socialization. And because of that, cognitive agency has two aspects that merit attention. The first of these is that learners usually become aware of their own cognitive agency at some point in early childhood. A young learner can—and almost always does—discover her own ability to produce novel information, and thereby come to develop a self-reflexive relationship between her intentions, desires, and her powers of constructive thought. At about the same time, a learner can also begin to witness the reliability of her own judgments about causal and statistical phenomena, while also becoming self-aware of the accuracy of some of her previously formed beliefs. Because of these different kinds of self-awareness, a learner can then take the mental steps involved in exercising her own cognitive agency simply by bringing judgment and imagination together.

This example allows us to make two important philosophical points. First, the concept of cognitive agency allows us to avoid one of the less desireable features of Piaget's earlier version of constructivism (Piaget 2001; Piaget and Duckworth 1970). Piaget was notoriously intellectualist in his treatment of children's development [cf. (Merleau-Ponty 2010)], largely ignoring the role that attachment and other affectional bonds play in maturation. But, perhaps because of this, he also frequently ignored the more specific issue of how the child's affective states, skills, and abilities interface with their (non-affective) cognitive abilities. We are attracted to the concept of cognitive agency because it overcomes this latter, more specific oversight—as a child's cognitive agency very likely depends on the stabilization and deployment of various different kinds of purely and partially affective capabilities, such as security, patience, curiosity, motivation, engagement, and so on. For many children, each of these affective states can become both a cause of *and* (reflexively) an object of the kind of self-awareness that can facilitate increases in cognitive agency.

Second, the example above also explains why cognitive agency will exhibit domain-specific and background-belief-dependent dynamics. In both cases, this is because cognitive agency can be influenced by a person's own self-directed beliefs. For example, someone who believes that they have a passion for marine mammals may have more stable cognitive agency when learning about fishes than someone without such a belief. Likewise, some domains may be more amenable to imaginative exploration by human minds than others—this may be a key difference between learning Euclidean versus non-Euclidean geometries. Or, someone who already knows a lot

about marine mammals may simply be better at constructing new marine-mammal-relevant ideas than someone without such knowledge.

The second aspect of cognitive agency that merits attention is its social aspect. All learning occurs in some social context, and so there is always the potential interplay between the features of a particular social context and the ability that learners have to exercise or inhibit their cognitive agency. Some environments may be extremely supportive of cognitive agency (science museums), while others may virtually suppress all cognitive agency (hospital waiting rooms). The interplay between cognitive agency, learning, and social contexts is another area where the normative and the psychological interact. But more on this topic of the social dimensions of cognitive agency in the next section, where we turn to the epistemological periphery of rational constructivism.

So, when it is realized, cognitive agency is expressed by such mental activities as tinkering with hypotheticals, random walk exploration, accidental discovery, and following crazy out-of-the-box instincts and hunches—all in such a way so that these activities are conducive, eventually, to the formation of beliefs that are probably true. Take for example Kekulé's report that his insight into the chemical structure of benzene came to him by way of a day-dream in which he saw an ouroboros (Brazier 1964, p. 334). The day-dream is a random walk, and the analogy between the snake and the ring structure of a set of molecules reflects another example of creative rationality. But recognizing that the hypothesis that benzene has a ring structure is probably true is an expression of Bayesian rationality. And so the underlying point here, which is nicely illustrated by examples like Kekulé's dream, is that the psychological processes that express cognitive agency can be quite complicated.

3.4 Philosophical Issues Related to Cognitive Agency

We have already observed that the relationship between cognitive agency and belief is reflexive, and this implies that a learner's beliefs can help or hinder the establishment of cognitive agency. In turn, this means that epistemological progress can be made by determining which beliefs—or which principles, if both believed and acted upon—are most conducive to establishing cognitive agency.

Yet, there is a limit to how far this line of investigation can be pursued if we are to use only philosophical methods. While it is true that this investigation involves looking for the beliefs and principles which can make a *normative* difference to a person's cognition (by promoting cognitive agency), we are also investigating a *causal* question, namely, do the same beliefs and principles make a *causal* difference to cognition (again, by promoting cognitive agency). Because the normative and the causal investigations are directed towards one and the same end, any meaningful progress will require some combination of philosophical and empirical methods.

There are, nevertheless, some philosophical observations that are worth briefly examining at this juncture, as they can clarify what might be involved in investigating the doxastic resources that are effective in promoting cognitive agency. First, we can introduce a term to refer to the set of beliefs, attitudes, principles (and whatever else) that helps to support cognitive agency. Let us call these various psychological entities a person's *latent epistemology*, because these are the beliefs (and whatever else) that do in fact make a difference to the maintenance of a person's cognitive agency, because at least some of these beliefs (and whatever else) will be epistemological in character, and

because these beliefs (and whatever else) are unlikely to be expressible in any organized or systematic fashion. A simple example of just such a belief might be a young child's belief that her imagination can be a source of *reliable* ideas about life, formed on the basis of playing with her sister in the woods.

Second, it is a mistake to think that, no matter how rich and systematic a person's latent epistemology can become, that this will be sufficient to stabilize and maintain cognitive agency. We mentioned above the important point that a learner's social context can dramatically impact how much cognitive agency she has. As a concrete illustration of this, consider the unfortunate situation in which a young child's parents make a point of discouraging curiosity and exploratory play at home. Despite these obstacles, this child may still come to believe in the power of her imagination to produce new thoughts and ideas, but nevertheless spend a great deal of her time in social contexts where the operative norms suppress her imagination, and thereby suppress her cognitive agency. So in addition to the role that a learner's own latent epistemology plays in maintaining cognitive agency, we also must take into account the role that social norms can play in doing the same.

Finally, it is worth saying a bit more about the metaphysics of the equilibrium that we believe defines cognitive agency in its ideal form. Indeed, please recall that we introduced the concept as a kind of normative rationality *only*, which gives us room to draw a distinction between real-world realizations of cognitive agency and what cognitive agency would look like in its ideal form. And in its ideal form, cognitive agency is whatever balance between (in turn) ideally rational constructive thinking (i.e. perfect creative rationality) and ideally rational Bayesian inference (i.e. perfect Bayesian rationality), as this balance will be extremely conducive to the formation of knowledge. So, there will likely be very few real-world cases of close-to-ideal cognitive agency. Nevertheless, whenever there is a productive interplay between well-functioning (but not perfectly rational) constructive thinking and well-functioning (but not perfectly rational) Bayesian inference, there is the possibility of cognitive agency, qua equilibrium, being realized. But at the same time, there is no reason to think that a learner's mind will "naturally" tend towards realizing cognitive agency. This is because our analysis has revealed, so far, that at least two kinds of exogenous influences are necessary: first, the learner's latent epistemology, and second, the epistemic norms that a learner encounters as she moves through all the different contexts and environments in which learning is possible. Having already briefly examined the relationship between a person's latent epistemology and their cognitive agency, we are free now to turn an examination of the second of the two influences.

4 The Epistemological Periphery: Two Principles for Learners

And in so doing, we shift our focus the epistemological periphery of rational constructivism, a shift which allows us to bring our attention to the question of what norms will support learning if our characterization of the epistemological core of rational constructivism is close enough to the truth. And the argument designed to highlight these norms is straightforward. It will be impossible to establish an equilibrium between creative rationality and Bayesian rationality if the social context of learning is not supportive of this outcome. To this end, we will argue here that there are two basic

principles that can function as schemas for thinking about how norms can be used to structure social practices that are conducive to learning.

4.1 Two Principles

The first principle is that learners have a prima facie right to develop and exercise their own creative rationality. And this right can be understood as a right to a kind of mental autonomy: the freedom to think actively and creatively, and the practical authority to demand this freedom in situations where learning by thinking may be possible. The right to creative rationality also creates for the learner the implied duty to practice, or otherwise work towards developing, their own creative rationality—something that the agentive aspects of creative rationality make psychologically possible.

Imputing to learners a right to creative rationality agency aligns the epistemology of rational constructivism with the extensive contemporary literature on play-based learning (Singer et al. 2006), and also earlier work on the social conditions of learning (Montessori 1978; Vygotsky 1980). For an example of the later, Montessori was keen to valorize what we are calling creative rationality. She takes note of the "creative instinct" and the "active potency" of a learner's mind, and the lessons she drew about how to structure a learner's environment—freedom of movement at school, the space and time and social support to "think for one's self", and also to try to learn, not by didaction, but by proactive physical and social interaction—provide a compelling real-world example of what realizing the right to cognitive agency looks like in practice. Some existing forms of education already appreciate the normative importance of cognitive agency.

The suggestion, then, is that a right to creative rationality secures for learners whatever is necessary to allow for constructive thinking to contribute as much as possible to the overall reliability of their beliefs. And recognizing this right, of course, places any number of different obligations on the people—parents, professors, teachers, friends, and so on—who are helping a learner learn, to, inter alia, not tell learners what to think and believe, but instead to cooperate with learners as they produce their own ideas and interpretations of the world around them. We stress that, as we argued previously, this is not irrational: the creative rationality of constructive thought must be assessed relative to the ideal of maximizing the amount of novel (and quite potentially mostly false) information that passes through a mind, for that will also, if the psychological details of rational constructivism are close enough to the truth, maximize the reliability of learning by forming beliefs according to some application of a Bayesian algorithm.

The second principle in rational constructivism's epistemological peripheral can be seen as a response to an important fact about learning that is usually ignored in contemporary philosophical epistemology: not all learners are epistemic agents; many are epistemic patients. Epistemic patients (most children, non-experts who are learning a new domain, recipients of some forms of testimony, and so on) do not have the ability to independently verify the truth or falsity of novel information, or determine with perfect reliability which people are trustworthy, and so on. Epistemic patients are the recipients of pedagogical attention and information from epistemic agents, but that is hardly the extent of the interaction between epistemic patients and agents. For, as it concerns cognitive agency specifically, we suggest that epistemic agents have a *prime*

facie obligation to help epistemic patients develop their own cognitive agency. The argument for this principle runs parallel to the argument for the prima facie right to creative rationality: individual learners are much more likely to be able to acquire knowledge if their social contexts provide them with the pedagogical scaffolds necessary to get to the point in their personal development at which they can become aware of, and subsequently learn how to exercise, their own cognitive agency.

Why don't we suggest that learners also have an epistemic right to Bayesian rationality? The short answer is that Bayesian rationality seems to be a kind of rationality that develops more or less independently and autonomously, unlike creative rationality and cognitive agency. It does not appear to depend, functionally speaking, on the psychological processes that implement conscious deliberation, choices, intention, motivation, and action; it does not seem that second-order beliefs can substantially alter the dynamics of Bayesian belief-evaluation. So, there does not appear to be an essentially agentive dimension to Bayesian rationality; as a matter of human psychology, it is not the kind of thing that a person can be responsible for. This means that a right to Bayesian rationality is no different from, for instance, a right to use one's eyes more or less as one pleases, or to feel as one feels most naturally as one moves through the day. Rights for these exist, but they are deeper moral and political rights—they do not find their home in epistemology.

4.2 Investigating a Person's Latent Epistemology

There are limits to how much can be said philosophically about which norms either do, or at least could, facilitate creative rationality and cognitive agency. The contexts in which people learn can change rapidly, and so too can the goals or outputs of different courses of learning; this means that largely a priori analysis will frequently be unable to shed much light on questions about which norms do in fact support (by helping to cause) learning. Nevertheless, the diversity of things that can impact learning reflects an opportunity for experimentation, simply because the more variation and internal interaction in a system, the more opportunity there is to learn about the system using experimental methods. Thus questions like "Which social norms support, inhibit, or amplify, cognitive agency?" are opportunities for new research programmes in psychology.

Indeed, returning briefly to an idea introduced several sections back allows us to expand upon this suggestion. We argued earlier that developmental and cognitive psychology have, and can continue to, each make an important contributions to answering the question: What epistemic concepts, beliefs, and principles do or should people have, in virtue of which they are able to engage in rational learning? We can add social psychology to the list of disciplines which, too, holds the promise of making important contributions to developing the fullest possible answer to this question. Of course, rational constructivism answers this question by, at bottom, claiming that the beliefs (in a person's latent epistemology) and norms (like the right to creative rationality, and the duty to promote cognitive agency) which stabilize or amplify cognitive agency are the ones which explain (at least in part) how it is that people do in fact engage in rational learning. But the deeper methodological point here is that other alternatives should be examined in detail. We noted in the introduction that rational constructivism is designed to be an alternative to popular nativist and empiricist

views in psychology; this means that there will almost certainly be nativist and empiricist answers, too, to question of what mental resources support and explain rational performance when learning.

5 Conclusion

The epistemological core of rational constructivism, then, is defined by the concepts of creative rationality, Bayesian rationality, and cognitive agency. The epistemological periphery of rational constructivism is constituted by at least two norms meant to promote creative rationality and cognitive agency respectively.

What new research ideas for psychology and cognitive science emerge from our efforts to work out the epistemology of rational constructivism? We can think of many. First, it is possible to empirically investigate the consequences of establishing a right to cognitive agency. To give a very simple example, you can compare the number of hypotheses that learners at schools which establish such a right can invent as explanations for some perfectly novel observation to the number of hypotheses that learners at school without such a right can invent. Second, the fact that the distinctions between creative rationality and Bayesian rationality, for instance, can be grounded in experimental evidence raises the intriguing question of whether there are other forms and levels of rationality—perhaps, for instance, social learning expresses its own distinctive type of rationality. Philosophers and psychologists both should find this question worthy of investigation because the forms of rationality that are realizable in the human mind are the point at which the normative and the causal interact, and so they define a nexus for epistemology and psychology. Finally, there is the causal question of what it means for the mind to generate novel information—for if we are right that fostering creativity, curiosity, counterfactual thinking, imagination, and similar cognitive abilities really does increase learning, there is a fascinating causal story to be told about how exactly the mind goes about the work of implementing these abilities [cf. (Lombrozo in press)].

We have one final philosophical observation to offer. Our approach to working out the epistemology of rational constructivism is, quite obviously, deeply influenced by Quine's call to "naturalize" epistemology. But we think that there are two important methodological points that are too frequently overlooked by both naturalists and their critiques in contemporary epistemology. Quine wrote:

> Epistemology, or something like it, simply falls into place as a chapter of psychology and hence of natural science. It studies a natural phenomenon, viz., a physical human subject. [...] The relation between the meager input and the torrential output is a relation that we are prompted to study for somewhat the same reasons that always prompted epistemology: namely, in order to see how evidence relates to theory, and in what ways one's theory of nature transcends any available evidence. [...] But a conspicuous difference between old epistemology and the epistemological enterprise in this new psychological setting is that we can now make free use of empirical psychology. (Quine 1969, pp. 82–84)

Quine is sometimes read as motivating two theses: first, that we can replace the analytical vocabulary of epistemology with either that of cognitive science, or perhaps

more realistically, a conjunction of the technical vocabularies of most of the cognitive sciences; and second, that the naturalistic project that Quine calls for will involve abandoning any concern for epistemic normativity. But we offer our analysis of the epistemology of rational constructivism as refutation of the thesis that either of these ideas is essential to epistemological naturalism. Our analysis demonstrates that the vocabulary of psychology will likely need to be extended to include any number of inherently epistemic and inherently normative concepts if we are to make sense of how the mind generates knowledge—even though it does not follow that these concepts (as is the case of, for example, the concept of creative rationality) are already in use in philosophical epistemology. Likewise, our analysis shows a patently naturalistic treatment of knowledge acquisition can lead to what for all appearances appear to be *sui generis* normative insights. And it is so much the better, we think, when any such insights can be grounded in compelling scientific evidence.

References

Bermudez, J. L. 2005. *Philosophy of Psychology: A Contemporary Introduction (Routledge Contemporary Introductions to Philosophy)* (1 edition). New York: Routledge.

Brazier, M.A.B. 1964. *Brain function*. Berkeley: University of California Press.

Carey, S. 1987. *Conceptual Change in Childhood*. Cambridge: MIT Press.

Carey, S. 2009. The Making of Abstract Concepts: A Case Study of Natural Number. In *The Making of Abstract Concepts*, ed. D. Marischal. New York: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199549221.003.013.

Carey, S. 2011. *The Origin of Concepts (Reprint edition)*. New York: Oxford University Press.

Carey, S. 2015. Why Theories of Concepts Should Not Ignore the Problem of Acquisition. In S. L. Eric Margolis (Ed.), *The Conceptual Mind: New Directions in the Study of Concepts.*

Cook, C., N. Goodman, and L. Schultz. 2011. Where science starts: Spontaneous experiments in preschoolers' exploratory play. *Cognition* 120: 341–349.

Danks, D. 2014. *Unifying the mind: Cognitive representations as graphical models*. Cambridge: MIT Press.

Davidson, D. 2001. *Inquiries Into Truth and Interpretation: Philosophical Essays*. New York: Oxford University Press.

Drayson, Z. 2012. The uses and abuses of the personal/subpersonal distinction. *Philosophical Perspectives. A Supplement to Nous* 26(1): 1–18.

Fodor, J.A. 2005. *Hume Variations*. London: Clarendon Press.

Frazier, B.N., S.A. Gelman, and H.M. Wellman. 2009. Preschoolers' search for explanatory information within adult–child conversation. *Child Development* 80: 1592–1611.

Gendler, T. 2000. *Thought experiment: On the powers and limits of imaginary cases*. New York: Garland Press.

Gentner, D., and C. Hoyos. 2017. Analogy and abstraction. *Cognitive Science* 9: 672–693.

Gopnik, A., & Bonawitz, E. 2015. Bayesian models of child development. *Wiley Interdisciplinary Reviews. Cognitive Science* 6(2): 75–86.

Gopnik, A., & Wellman, H. M. 2012. Reconstructing constructivism: Causal models, Bayesian learning mechanisms, and the theory theory. *Psychological Bulletin* 138(6): 1085–1108.

Gopnik, A., L. Schulz, J.B. Tenenbaum, T.L. Griffiths, and S. Niyogi. 2007. *Causal learning*. London: Oxford University Press.

Griffiths, T. L., Kemp, C., & Tenenbaum, J. B. 2008. Bayesian models of cognition. In *The Cambridge Handbook of Computational Psychology*. London: Cambridge University Press.

Griffiths, T.L., N. Chater, C. Kemp, A. Perfors, and J.B. Tenenbaum. 2010. Probabilistic models of cognition: Exploring representations and inductive biases. *Trends in Cognitive Sciences* 14: 357–364.

Haith, M. M. 1998. Who put the cog in infant cognition? Is rich interpretation too costly? *Infant Behavior & Development* 21(2): 167–179.

Harris, P.L. 2000. *The work of the imagination*. New York: Blackwell Publishing.

Jones, M., & Love, B. C. 2011. Bayesian Fundamentalism or Enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition. *The Behavioral and Brain Sciences 34*(4): 169–188 disuccsion 188–231.

Lane, J. D., Ronfard, S., Francioli, S. P., & Harris, P. L. 2016. Children's imagination and belief: Prone to flights of fancy or grounded in reality? *Cognition* 152, 127–140.

Lombrozo, T. 2012. Explanation and abductive inference. *Oxford Handbook of Thinking and Reasoning*. New York: Oxford University Press.

Lombrozo, T. In press. "Learning by thinking" in science and in everyday life. In *The Scientific Imagination*, ed. P. Godfrey-Smith and A. Levy. New York: Oxford University Press.

Marr, D. 2010. *Vision: A Computational Investigation Into the Human Representation and Processing of Visual Information*. Cambridge: MIT Press.

Merleau-Ponty, M. 2010. *Child Psychology and Pedagogy: The Sorbonne Lectures 1949-1952*. Chicago: Northwestern University Press.

Montessori, M. 1978. *The Secret of Childhood*. Hyderabad: Orient Blackswan.

Oaksford, M., & Chater, N. 2009. Précis of bayesian rationality: The probabilistic approach to human reasoning. *The Behavioral and Brain Sciences* 32(1): 69–84 discussion 85–120.

Otto, L., & Rusanen, A. 2011. Turing machines and causal mechanisms in cognitive science. In *Causality in the Sciences*. Oxford: Oxford University Press.

Piaget, J. 1954. *The construction of reality in the child*. New York: Basic Books.

Piaget, J. 2001. *The Language and Thought of the Child (Routledge Classics) (Volume 52)* (1 edition). New York: Routledge.

Piaget, J., and E. Duckworth. 1970. Genetic Epistemology. *The American Behavioral Scientist* 13 (3): 459–480.

Quine, W.V.O. 1960. *Word and object*. Cambridge: MIT Press.

Quine, W.V.O. 1969. Epistemology Naturalized. In *Ontological Relativity and Other Essays*, ed. W.V.O. Quine. New York: Columbia University Press.

Rescorla, M. 2013. Rationality as a constitutive ideal. In A Companion to Donald Davidson (pp. 472–488). New York: Wiley.

Sim, Z. L., & Xu, F. 2017. Learning higher-order generalizations through free play: Evidence from 2- and 3-year-old children. *Developmental Psychology* 53(4): 642–651.

Singer, D.G., R.M. Golinkoff, and K. Hirsh-Pasek. 2006. *Play = Learning: How Play Motivates and Enhances Children's Cognitive and Social-Emotional Growth*. New York: Oxford University Press.

Spelke, E. S. 1998. Nativism, empiricism, and the origins of knowledge. *Infant Behavior & Development* 21(2): 181–200.

Spelke, E.S. 2003. What makes us smart? Core knowledge and natural language. In *Language in mind: Advances in the study of language and thought*, ed. Dedre Getner and Susan Goldin-Meadow, 277–311. Cambridge: MIT Press.

Stanovich, K.E. 2012. On the distinction between rationality and intelligence: Implications for understanding individual differences in reasoning. In *The Oxford Handbook of Thinking and Reasoning*, ed. Keith J. Holyoak and Robert G. Morrison. New York: Oxford University Press.

Stanovich, K.E., and R.F. West. 1998. Individual differences in rational thought. *Journal of Experimental Psychology. General 127* (2): 161.

Stich, S. P. 1978. Beliefs and subdoxastic states. *Philosophy of Science* 45(4): 499–518.

Taylor, M. 2013. *The Oxford handbook of the development of imagination*. New York: Oxford University Press.

Tenenbaum, J.B., C. Kemp, T.L. Griffiths, and N.D. Goodman. 2011. How to grow a mind: Statistics, structure, and abstraction. *Science* 331: 1279–1285.

Vygotsky, L.S. 1980. *Mind in Society: The Development of Higher Psychological Processes*. Cambridge: Harvard University Press.

Walker, C. M., Gopnik, A., & Ganea, P. A. 2015. Learning to learn from stories: children's developing sensitivity to the causal structure of fictional worlds. *Child Development* 86(1): 310–318.

Waxman, S. R., & Markow, D. B. 1995. Words as invitations to form categories: Evidence from 12- to 13-month-old infants. *Cognitive Psychology* 29(3): 257–302.

Weisberg, D. S. 2016. How fictional worlds are created. *Philosophy Compass* 11(8): 462–470.

Welder, A.N., and S.A. Graham. 2001. The influence of shape similarity and shared labels on infants' inductive inferences about nonobvious object properties. *Child Development* 72: 1653–1673.

Xu, F. 1997. From Lot's wife to a pillar of salt: Evidence for *physical object* as a sortal concept. *Mind and Language* 12: 365–392.

Xu, F. 1999. Object individuation and object identity in infancy: The role of spatiotemporal information, object property information, and language. *Acta Psychologica* 102(2-3): 113–136.
Xu, F. 2002. The role of language in acquiring object kind concepts in infancy. *Cognition* 85: 223–250.
Xu, F. 2007. Rational statistical inference and cognitive development. In *The innate mind: Foundations and the future*, ed. P. Carruthers, S. Laurence, and S. Stich, vol. 3, 199–215. New York: Oxford University Press.
Xu, F. 2016. Preliminary thoughts on a rational constructivist approach to cognitive development. In *Core knowledge and conceptual change*, ed. D. Barner and A. Baron. New York: Oxford University Press.
Xu, F. and Griffiths, T. 2011. Probabilistic models of cognitive development: Towards a rational constructivist approach to the study of learning and development. *Cognition* 120(3): 299–301.
Xu, F., & Kushnir, T. 2012. *Rational constructivism in cognitive development* (Vol. 43). London: Academic Press, Elsevier.
Xu, F., & Kushnir, T. 2013. Infants are rational constructivist learners. *Current Directions in Psychological Science* 22(1): 28–32.
Xu, F., & Tenenbaum, J. 2007. Word learning as Bayesian inference. *Psychological Review* 114(2): 245–272.