ARTICLE IN PRESS

Evolution and Human Behavior xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Evolution and Human Behavior

journal homepage: www.elsevier.com/locate/ens

# Psychological essentialism from first principles☆

Tadeg Quillien*

*Center for Evolutionary Psychology, Department of Psychological and Brain Sciences, University of California Santa Barbara, USA*

ARTICLE INFO

*Keywords:*
Essentialism
Concepts
Categorization
Induction
Causality
Object files

ABSTRACT

Cognitive scientists have documented the existence of "essentialist" intuitions in humans: from a very early age, we assume that things have deep unobserved properties that make them what they are. I provide a sketch of an adaptationist explanation of psychological essentialism, arguing that these intuitions are the unsurprising output of adaptations for inductive inference. Variations on this insight have previously been used mostly as after-the-fact speculations, yet theories of adaptive function should ideally have a primary role in informing psychological research. Here I propose that viewing essentialist intuitions through an adaptationist lens has implications for some widespread assumptions about the phenomenon. Notably, researchers' focus on "higher-level" processes like categorization has led them to assume that essentialism is restricted to a few cognitive processes, but the ubiquity of inductive inference problems in cognition suggests otherwise. Additionally, because essentialist intuitions are the output of mechanisms solving related but distinct inference problems, it is unlikely that a single mechanistic theory can account for them all.

Cosmides and Tooby (1994) have argued that progress in cognitive science has been limited by a failure to think in rigorous adaptationist and computational terms. This is because "intuition systematically blinds us to the full universe of problems our minds spontaneously solve, restricting our attention instead to a minute class of unrepresentative 'high-level' problems" (p.41). Here I propose that research on "psychological essentialism"[1] provides a good case study of this problem.

Starting in the late 1980's, cognitive scientists have documented an impressive array of 'essentialist' intuitions that humans have from a very early age (Gelman, 2003). Essentialism is a constellation of phenomena, but its main components can be roughly summarized as the following. (1) Human concepts are not simple lists of features but involve the representation of yet-to-be-discovered entities: people know that something can look like gold but be something else, because their concept of gold is not a simple list of perceptual features (like being yellow and shiny): instead they think that whether something counts as gold can also be determined by some (possibly yet undiscovered) non-obvious properties. (2) Causal relationships play an important role in the way concepts are structured: people assume that some features of an entity are causally related to each other, and this causal information plays a crucial role in people's categorization decisions, in that causally deep features are more important. For instance, people may assume that something in the atomic structure of gold *causes* gold to be yellow and shiny, and that having this particular atomic structure is more important to what it means to be gold than being yellow and shiny. (3) People have strong domain-specific assumptions about categories. For instance, they understand that the "insides" of an animal are important in determining the properties of this animal (Gelman & Wellman, 1991; Keil, 1989); by contrast they do not rely on this principle when reasoning about artifacts.

The research program on psychological essentialism is mostly concerned with "high-level" cognitive processes such as categorization. It has also mostly raised questions of adaptive function as an afterthought, if at all. For instance, the most authoritative book on essentialism (Gelman, 2003), asks "why do we essentialize?" only in the very last chapter, in a section dedicated to speculations.

Yet without a theory of adaptive function to guide one's investigation, discovering the cognitive mechanisms that underlie a set of phenomena is like looking for a needle in a haystack (Cosmides & Tooby, 1994). One may get a rough idea of where to start looking by consulting one's common-sense intuitions about the mind, but those are a poor

☆ This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

* Department of Psychological and Brain Sciences, University of California Santa Barbara, Santa Barbara, CA 93106, USA.
*E-mail address:* tadeg.quillien@psych.ucsb.edu.

[1] "Psychological Essentialism" can refer to two distinct things: one is a set of empirical findings, the other is a specific theory of cognitive architecture that is meant to explain these findings. In order to distinguish these, I will use "essentialism" to refer to the empirical findings, and "Psychological Essentialism Theory" (PET) in order to refer to a specific psychological theory that is meant to explain the findings. Throughout the paper I am mostly concerned with essentialism (the phenomenon itself), although I discuss PET toward the end.

https://doi.org/10.1016/j.evolhumbehav.2018.07.003
Received 10 October 2017; Received in revised form 18 July 2018; Accepted 18 July 2018
1090-5138/ 

guide to the actual structure of cognition, notably because most of the computations our brain performs are outside of conscious awareness. Instead, investigation of a cognitive system ideally starts by specifying the information-processing problem this system is designed to solve. Then, one proceeds to find a good solution to this problem (a *computational theory*, Marr, 1982). The structure of the cognitive system one studies is likely to implement an approximation to such a solution.

Marr uses the example of a cash register to illustrate the concept of a computational theory. To understand how a cash register works, one has to understand the problem it is designed to solve. The machine has to compute the total price a customer must pay for the contents of his basket. A number of requirements follow from this problem: if I buy nothing, it should cost me nothing; if I buy nothing and then something, it should cost me as much as if I had bought just the something; the order in which the goods are presented to the cashier should not affect the total I have to pay; etc. The complete list of requirement turns out to be equivalent to the definition of the arithmetic operation of addition: therefore, we can summarize the design requirements of a cash register by saying that it has to be able to perform additions.

The computational theory of an information-processing problem is agnostic about the way the computation is performed. It specifies that the cash register has to perform addition, not whether it should encode the numbers in base 2 or base 10. This latter kind of question belongs to what Marr calls the *algorithmic level* of analysis. This level describes the format in which the relevant information is represented, and the algorithms that operate over these representations. Importantly, for a given computational theory, there is an infinity of possible algorithms that can perform the specified computations: for instance, the same addition can be performed using binary, decimal, hexadecimal numbers, etc.

Research on essentialism has almost exclusively focused on the algorithmic level of analysis: researchers have sought to explain their data using theories of the structure of people's representations and the algorithms that operate over these representations. The most popular such theory, Psychological Essentialism Theory (PET), claims that people represent categories as having an essence, a feature whose possession is necessary and sufficient for being the thing in question (Gelman, 2003; Medin & Ortony, 1989). This focus on the algorithmic level makes PET a *proximal* explanation.

By contrast, here I want to explore *ultimate* explanations to essentialism. I provide a sketch of an adaptationist framework within which to understand the empirical findings on essentialism, mostly at the computational level of analysis. I show that, to a large extent, this framework validates the main conclusions of the research program: for instance, on an evolutionary basis we strongly expect essentialism to be a universal, reliably-developing phenomenon (Gelman, 2003) as opposed to a cultural construction (Fodor, 1998). On the other hand, I propose that the neglect of the computational level of analysis has forced researchers to ground their work on unreliable folk-psychological concepts and intuitions. Because most research on essentialism is restricted to relatively "high-level" processes, it is commonly assumed that essentialism is a superficial add-on to an otherwise non-essentialist cognition (for instance, Fodor (1998), views essentialism as a cultural innovation; Gelman (2003) doubts that it occurs in non-human animals). Also, current work tends to assume that the various components of essentialist intuitions can be explained by some underlying common structure or cognitive bias (Cimpian & Salomon, 2014; Gelman, 2013; Newman & Knobe, 2018) – for instance, that the same general representational structure explains both why people think of things as having undiscovered features and why they spontaneously intuit that animals inherit some properties from their parents (Ahn et al., 2001). I will argue that, from a computational point of view, these assumptions are not necessarily warranted.

I first identify a class of information-processing problems that constitute adaptive challenges for which essentialism may be relevant (Section 1). In later sections I describe some design requirements for cognitive systems that would have to solve these information-processing problems. These systems have to be non-descriptivist (Section 2), be able to represent causal relations (Section 3) and to use domain-specific knowledge (Section 4). I show that essentialist intuitions that have been documented in humans follow directly from these design requirements.

In this light, I then propose that it is not necessary to posit the existence of a single representational structure to explain essentialist intuitions (Section 6), and that theories that posit such a single structure are implausible (Section 7).

## 1. Identifying a class of information-processing problems

Essentialist intuitions are intuitions about category membership. There is a general consensus that the main function of categorization is to promote inductive inferences (Anderson, 1991; Barrett, 2001; Gelman & Coley, 1991). Therefore, it is very likely that essentialism is adaptive because it helps people make better inductive inferences (Barrett, 2001; Gelman, 2003; Gelman & Coley, 1991). The idea is that, because the world does have a complex causal structure, cognitive mechanisms that assume such a complex structure when constructing and using categories have an epistemic edge over those that do not. In the next paragraphs, I elaborate on this idea by proposing that the set of information-processing problems for which essentialist cognition is adaptive is larger than typically recognized.

Cognitive scientists say that a category like "birds" has a "rich inductive potential" to mean that because birds have a lot of features in common with each other, something that one learns about a particular bird can easily be generalized to most other birds. Categories like "birds", "gold", or "trees" that have a very rich inductive potential are typically called "natural kinds" to highlight the fact that they are not arbitrary groupings. Researchers have proposed that essentialism is useful because it enables category-based inference, most notably for natural kinds (Barrett, 2001; Gelman, 2003; Gelman & Coley, 1991). However, terms like "natural kinds" and "categories" are not very well defined from a computational point of view. For instance, if natural kinds are those kinds that have a rich inductive potential, then there is no clear delimitation between what counts as a natural kind and what does not: instead there is a continuum between kinds that are very natural and kinds that are very unnatural, with "tigers" and "white things" at each extreme, and kinds such as "chair" somewhere in the middle (Markman, 1989; Millikan, 1998). "Category" is also a problematic concept, because intuitively the term refers to a set of several items (like "birds", or "US presidents"), rather than one single individual (like "this bird" or "John Kennedy"). But even a single individual can have a rich inductive potential: because people do not change a lot over time, something one learns about a person can be predictive in future encounters with that person. Essentialist intuitions can apply to individual concepts, as when people think that the original Mona Lisa is more valuable than a perfectly identical copy (Gelman, 2013). Therefore, it is convenient to have a term that refers to anything that has *some* inductive potential, whether a category or an individual.

To this end, Millikan (1998) introduces the concept of *substance*. A substance is simply an entity "about which it is possible to learn from one encounter something about what to expect on other encounters" (p.57). For instance, the kind "birds" is a substance because what one learns about a particular bird might be generalized to other birds. The individual "Jack" is also a substance because what we learn on one encounter with Jack is predictive in future encounters with the person. The use of extra jargon in science is usually bad, but adopting the concept of substance is justified here, because the ontology that is provided by common sense is somewhat unhelpful.[2] Although kinds

[2] The use of a strange term like "substance" instead of more familiar terms like "category" is not meant to imply that the world is somehow fundamentally

and individuals are different in a lot of ways, having a term encompassing both highlights the fact that their representations perform a very similar computational function. When I enter a room and see someone standing there, if I recognize this person as none other than my friend Jack, I am able to make a lot of new inferences about the-person-in-the-room (if I know that Jack is shy, I can predict the behavior of the-person-in-the-room). Similarly, if I see an animal on a tree branch, if I recognize this animal as a bird, I can make a lot of new inferences about the-animal-on-the-branch (I can predict that it can fly). In both cases we use the representation of a substance for inferential purposes.

The world contains many substances, which by their very nature can be exploited for information-gathering purposes by organisms (Millikan, 1998). One fundamental adaptive problem for any organism is to construct and maintain good representations of fitness-relevant substances, and use these representations to generate inferences. This is an extremely general problem. Organisms have to make inferences from sparse data all the time, and those are notoriously difficult (Chomsky, 1959; Markman, 1989; Tooby & Cosmides, 1992). They require some background knowledge that can be provided by a representation of the relevant substances. While our common-sense ontology makes appealing the notion that there is such a discrete thing as a problem of category-based inference, or of natural-kind-based inference, from a computational point of view, any substance that has *some* inductive power presents information-gathering opportunities, and exploiting these opportunities is an adaptive problem. While Gelman (2013, p.460) recognizes that "all higher thought entails going beyond the information given", inferences from sparse data are equally characteristic of "lower" mental processes like vision (Helmholtz, 1925; Knill & Richards, 1996; Marr, 1982). Focusing on "higher" problems unnecessarily restricts one's attention to an unrepresentative subset of the inference problems our minds routinely have to solve.

There is a wide variety of design requirements for a cognitive system that has to perform substance-based inference. In the next sections I lay out some of these design requirements, and show that they can explain essentialist intuitions in humans.

## 2. Non-descriptivism

One possible design for substance representation systems is for the representation to just be a list of the features that the substance has. Traditionally this idea has been associated with theories of concepts such as classical and prototype theories (see Smith & Medin, 1981, for review). The idea is that people's mind contains a list of features that birds do and do not have, and in order to decide whether an X is a bird, they just check the features of that X against the list. If there is a sufficient match between the features of X and the list of bird-typical features, then X is recognized as a bird. The list of features, together with a procedure that specifies what constitutes a match to that list, is all there is to the representation of "birds". This procedure can be very sophisticated (e.g., it can be probabilistic); nonetheless, the list of features and the associated identification procedure are all there is to the representation.

The doctrine that representations are just lists of features is known as *descriptivism* (see Recanati, 2012). From a computational point of view, one does not expect evolved organisms to have descriptivist representations, because these have several weaknesses.

First, constructing the correct representation of a substance is an inference problem. People do not have innate knowledge about birds; they must learn that "birds" are a meaningful category, and learn how to best identify birds, on the basis of limited data. Especially at the early stages of this inference process, the tentative list of features that they have collected is very unlikely to be the correct one. A child might have an early representation of birds that includes bats and pterodactyls but excludes ostriches, and then gradually improve this representation over time, by learning more about the features that birds do and do not possess (e.g. learning that all birds have feathers). Because inferring the right representation of "birds" is an empirical problem, there is no a priori way for an organism to know when its current list of features is perfectly accurate.

A related problem is the fact that substances can change over time. That Jack is wearing a blue t-shirt today does not mean that he will wear one tomorrow. Our representation of Jack should include the information that Jack is wearing a blue T-shirt, but the slot in our representation that says "blue T-shirt" should be allowed to be overwritten, because people can change clothes while retaining their other characteristics. More strikingly, the Ship of Theseus has its pieces continuously replaced: one can imagine that a version of the Ship ten years from today would have no single feature in common with the one we see now, but assuming that the changes are gradual enough, the Ship is a valid substance, because most of what one knows about today's version of the Ship can be generalized to tomorrow's version. In our representation of the Ship, no feature should be immune from replacement.

Therefore, well-designed cognitive programs for representing substances are expected to include procedures for modifying the list of substance-typical features, in order to make it more accurate, or keep it synchronized with the actual state of the world. The system should also have an estimate of how well the substance is currently represented, such that it knows how to weigh new information about it (e.g., a botanist should be more reluctant to change his representation of what a tree is than a child should be, because he knows that he already has a relatively good grasp on the matter). Descriptivist representations, as defined above, contain no information about the uncertainty that the system has about the representation's accuracy, and do not allow for the representation to be improved over time, or for parts of the representation to be overwritten to keep track of a changing world, because they do not have a procedure for modifying the list of features (indeed, modifying the list of features would amount to creating a new concept entirely). Therefore they are not a plausible design for cognitive systems of actual organisms (see also Millikan, 1998).

A major impetus for the claim that humans are naïve essentialists is the fact that we have non-descriptivist representations. In particular, research has shown that representations are not simple lists of perceptual features, even in children. Gelman and Markman (1986, 1987) showed that preschoolers are able to generalize non-obvious properties across category members on the basis of category labels, and that these labels override perceptual similarity. For instance, 4-year-olds were shown a dolphin, a shark, and a tropical fish, and told that a shark and a tropical fish are both fishes. When told that the shark stays underwater to breathe, children preferentially extended this property to the tropical fish rather than the dolphin, despite the fact that the dolphin was more similar in appearance (see also Gelman & Coley, 1990; Welder & Graham, 2001). Gelman and Wellman (1991) find that preschoolers consider non-visible properties ("insides") of some objects to be more important to category membership than their visible "outsides" in determining category membership. Non-descriptivism also characterizes children intuitions about stability over transformation: 7-year-olds consider that a raccoon made to look like a skunk is still a raccoon (Keil, 1989; see Rips, 1989 for similar results in adults). Finally, spatio-temporal continuity, more than physical features, plays a major role in people's intuitions about persistent identity over time (Gelman, 2013; Hall, 1998). These empirical findings are unsurprising, given the recognition that descriptivism does not meet the design requirements of cognitive systems for substance-based inference.

(*footnote continued*)
different than what people think, or that the familiar terms do not describe real and important things. Describing the world in terms of substances is merely more helpful in the context of this paper, because of its emphasis on the computational level of analysis.

## 3. Causal inference

Some substances get their rich inductive potential from the fact that all their features have a common cause: most of the properties of gold can be explained by its chemical structure (having atomic number 79). More generally, the features of a substance are very often linked by causal relationships: birds can fly because they have wings, and have wings because their DNA codes for the development of wings. Representing the correlation between these features is adaptive, but representing them as being the symptom of causal relationships is even better. Suppose a person's representation of "birds" includes the information that there is a correlation between having wings and being able to fly (such that birds with no wings generally cannot fly, and vice-versa). The observed correlation can be explained by several competing causal models: having wings might cause the ability to fly, the ability to fly might cause the possession of wings, or a third variable might cause both. Although equally consistent with the observed correlation, each of these models also makes different predictions. On the "flying causes wings" causal model, cutting a birds' wings should not deprive it of the ability to fly, but it should do so on the "wings cause flying" model. Therefore, getting the right causal model is important, and we would like our cognitive systems for substance-based inference to be able to adequately extract and use causal information.

Causal relations in the world are not transparently visible, and so organisms must solve a 'causal inverse problem' (Gopnik et al., 2004). They have to infer causal facts from limited data – in the same spirit that the visual system must solve a 'spatial inverse' problem of inferring the 3-D structure of the world from the information reaching the 2-D surface of the retina. They can solve the causal inverse problem by performing direct interventions on their environment, and by integrating observational data about correlations with prior assumptions about the way the world works (e.g., causes usually precede effects, diseases cause symptoms rather than vice-versa, etc.). By doing so, they can construct causal maps of the world, internal representations of the causal relationships between variables from which predictions can be made (Gopnik et al., 2004).

Research in artificial intelligence provides computational theories of the problems involved in the acquisition and use of causal knowledge, thanks to formalisms such as Bayesian causal networks (Pearl, 2000; Glymour, 2001; Griffiths & Tenenbaum, 2005). Briefly, a Bayesian causal network is a formal model of the causal relations between variables, which uses probability theory and basic principles about causality to make appropriate inferences. For instance, a causal network can be used to represent the fact that either rain or a water sprinkler can make the pavement wet; and upon learning that the pavement is wet and that the water sprinkler is off, the network can infer that it has been raining (Pearl, 2000).

Human cognition turns out to be accurately described by these computational models – for instance, children seem to extract causal information in a way that is consistent with the prescriptions of Bayesian theories of causal learning (Gopnik et al., 2004; Schulz, 2012). Computational theories of causal cognition also shed light on results in the essentialist literature that involve the representation of the causal structure of categories. Medin and Ortony (1989) have argued that one important component of essentialist thinking is the fact that people think of categories as having a "core" - a small set of features which are crucial for category membership, and tend to have causal power in producing the other features of category members. Ahn, Kim, Lassaline, and Dennis (2000) have shown that people think that the deeper a feature lies in a causal chain, the more central it is to category membership. Given causal information in the form of a chain such as Bird DNA → Wings → Flight, they think that having bird DNA is more central to what it means to be a bird than having wings, which is itself more central than being able to fly. This effect of causal depth on feature centrality has been shown formally to be a requirement of computational theories of causal representation, given certain minimal assumptions (Rehder, 2007), thus providing an ultimate explanation for the phenomenon.

Computational theories of causal inference also bring an additional piece of explanation to empirical results such as the finding by Keil (1989), that children are not fooled by surface transformations, and judge that a raccoon dressed up as a skunk is still a raccoon. We saw in the previous section that an organism should not rely exclusively on surface features when determining substance membership. But why is the child not compelled by the change in surface features to infer that the other, non-visible features of the animal were also changed? The answer is simple if we assume that the child maintains an internal map of the causal relationships between category features. Because the child is aware that the cause of the changes in the surface features of the animal is exogenous – they are the responsibility of the experimenter – the new surface features do not prompt the child to draw new inferences about the causally deep properties of the animal. As we just saw, the latter are central to substance membership, and therefore the child has no reason to re-categorize the animal.

## 4. Domain-specific knowledge

Inductive inference is a difficult problem because it requires *inductive constraints*, which help an organism select the best one among the infinity of possible inferences that it could draw from a given set of data. Some of these constraints are very general: for instance, good inference should be consistent with the laws of probability; also, as argued above, inductive inference systems should not in general be descriptivist. However, inference also relies on relevant background knowledge, which will vary from one domain to the next. Therefore, on functional grounds one does not expect inductive inference to be governed exclusively by a rigid single set of principles; inference systems should exhibit some degree of specialization to their domains (Boyer & Barrett, 2015).

To give an example, birds are animals, and therefore specialized knowledge about animals is helpful in order to solve the difficult problem of inferring the correct representation of "birds" from sparse data. An organism correctly guessing that birds are animals can construct a representation of birds that inherits all the general properties of animals: animals have to breathe, eat, reproduce, can move, etc., and by extension birds do as well. Knowledge about animals also contains abstract information about the way an animal species is structured, like the fact that species membership can be inherited genealogically, so that the offspring of a bird will necessarily also be a bird. It also contains information about the features that can be validly generalized from a single substance members to all other members. For instance, the fact that an ostrich lays eggs can be generalized to all birds, but its size cannot.

Note that arguing that domain-specificity is essential to successful inference is not the same as arguing that domain-specific knowledge is genetically encoded. Some researchers have argued that domain-specific knowledge could in principle be learned, thanks to sophisticated statistical learning algorithms that construct hierarchical models of the world, where knowledge about one level (e.g. knowledge about birds; knowledge about animals) constrains and guides learning at the other levels. These algorithms are theoretically capable of acquiring even very abstract principles (see Kemp & Tenenbaum, 2008; Tenenbaum, Griffiths, & Niyogi, 2007; Tenenbaum, Kemp, Griffiths, & Goodman, 2011). Research in this tradition is nonetheless committed to the idea that domain-specific knowledge is essential for successful inference (Tenenbaum et al., 2011). Therefore, however much of mental content is actually genetically encoded, one expects the mind, on functional grounds, to eventually develop specialized inference systems.

It is indeed the case that people's intuitions about entities in a domain are guided by very rich background knowledge about that domain (what cognitive scientists call intuitive theories). Most of the evidence for this claim has come from studies of people's intuitions about the biological domain. Humans seem to possess universal, early-developing intuitions about the biological world (Atran, 1998; Inagaki & Hatano, 2006), and they use these intuitions to guide the inferences they make about living things.

Many essentialist intuitions involve such domain-specific knowledge about the biological world. For instance, preschoolers grant more importance to biological heredity than environmental factors when predicting the features of biological entities. Gelman and Wellman (1991) showed that 4-year-olds consider that an apple seed put in a flower pot will grow into an apple tree, and a baby rabbit raised by monkeys will grow up to prefer carrots to bananas. Hirschfeld (1995) similarly found that preschoolers predict the skin color of an adopted child to match that of its biological rather than its adoptive parents. Children are also sensitive to the privileged role of internal features in guiding inductive generalization about animals (Gelman & Wellman, 1991), even in infancy (Newman, Herrmann, Wynn, & Keil, 2008; Setoh, Wu, Baillargeon, & Gelman, 2013).

Assuming the existence of this abstract domain-specific knowledge completes the explanation of Keil's (1989) transformation experiments. How do children know that raccoons get most of their inductive potential from internal, as opposed to superficial, features? It is hazardous for them to make the blanket assumption that every substance gets its inductive potential from non-visible features, because there are substances for which this is false: there is nothing hidden within a birdfeeder that is responsible for all the artifact's properties. Rather, they rely on abstract knowledge they have about living things – knowledge that might be innate, or learnt over the years thanks to powerful mechanisms for hierarchical probabilistic inference. This body of biological knowledge contains the abstract principle that animals get most of their inductive potential from internal features, and the child's judgments are guided by this principle.

Keil's data suggest that domain-specific principles are indeed applied where they are relevant. When reasoning about artifacts, children do not use principles that apply to the biological world. Children were willing to accept that external transformation could change the category membership of an artifact, such that a coffeepot made to look like a birdfeeder is now a birdfeeder; this result contrasts with their intuitions about animals. Note that this finding does not imply that people are not essentialist about artifacts; they merely use different criteria, like creator's intent (Bloom, 1996) to infer substance membership in that domain (when an artifact is transformed by accident, people typically do not re-categorize it).

As Strevens (2001) argued, postulating a domain-general essentialist bias is an incomplete explanation for most essentialist intuitions – the latter often cannot be made sense of without understanding the nature of the domain-specific knowledge people use to guide their inferences (see also Barrett, 2001; Boyer, 1998, 2000).

## 5. Interim summary

Just as listing the design requirements for a cash register yields the definition of the arithmetic operation of addition, listing some of the common design requirements for systems of substance-based inference yields the main characteristics of essentialist intuitions that have been documented in humans. Assuming that natural selection creates organisms with well-designed cognitive adaptations for substance-based inference, it is not surprising that humans have universal, reliably-developing essentialist intuitions. An organism unable to dynamically update the content of his substance representations, unable to represent the causal relationships between substance features, and unable to draw on abstract domain-specific principles to guide his inferences would be doomed to reproductive oblivion. From this adaptationist point of view, it is very unlikely that essentialism is only a recent cultural construction (Fodor, 1998) or a consequence of language (Carey, 1996).

The view that essentialism is the symptom of a suite of cognitive adaptations for problems of substance-based inference has broader implications. If essentialist intuitions can be explained at the computational level of analysis, a single unifying algorithmic-level account of these intuitions may be unnecessary. Such a unifying theory might even be impossible, because one expects inference systems to be specialized.

By contrast, essentialist intuitions are often explained by Psychological Essentialism Theory (PET), a unifying algorithmic-level theory that aims at explaining all essentialist intuitions as falling out of the same representational structure. In the next sections, I therefore suggest that the adaptationist account warrants some skepticism about the explanatory scope of PET.

## 6. A unifying algorithmic theory of essentialist intuitions may be unnecessary

Above I have given a sketch of an explanation of essentialist intuitions at the computational level of analysis. This raises the following question: to what extent can one explain essentialism at the algorithmic level? Recall that in Marr's scheme (Marr, 1982) the algorithmic level specifies the kinds of algorithms that perform the computations, as well as the representational format these algorithms use.

Essentialist intuitions are often explained at the algorithmic level by PET, a theory about the structure of some mental representations. Before describing PET in more detail, it is worth asking whether the current computational account can, even in principle, have implications for the theory. Computational and algorithmic theories belong to different levels of explanations, and therefore they are not directly competing with each other. For instance, many different algorithmic theories of a cognitive system can be consistent with a given computational theory. If one could not open the insides of a cash register, one could still derive a theory of what the cash register does by observing its behavior, but one could not say on this basis whether its circuits are representing numbers in base 2 or base 10, or whether it uses serial or parallel computing.

Although theories at different levels of explanation do no directly compete with each other, they can still have implications for one another. Consider the following example, which uses Tinbergen's (1963) typology of explanations in biology. One can explain why bats have wings by pointing to some embryological process (say, embryological process X) that reliably results in the development of wings in bats: this is an ontogenetic explanation. One can also show that wings enable bats to fly, which has adaptive advantages: this is a functional explanation. Although the two explanations do not directly compete with each other, the functional explanation has implications for the *scope* of the ontogenetic explanation. If one finds that wings have an adaptive function in all taxa in which they exist, then they may have evolved via convergent evolution; if so, the only explanation of wings that is general enough to apply to all taxa may be the functional explanation. That is, maybe no unique embryological process can account for the existence of wings in all these taxa. Embryological process X may still be a valid ontogenetic explanation of bat wings, but we would now be cautious about extending this explanation to butterflies. By contrast, if it appears that wings have no adaptive function, then the existence of a single, conserved embryological process is a more likely explanation to the existence of wings across these taxa.[3]

Similarly, one can argue that, given the proposal that essentialist intuitions follow from basic design requirements of mechanisms for substance-based inference, one might not need to assume that all essentialist intuitions are explained by the same algorithmic theory. That

[3] Note that I am using this analogy to make a general point about how, in science, explanations at different levels can have implications for one another. I am not making a point about convergent evolution in particular. Notably, when I argue below that the computational account has implications for the scope of an algorithmic account of essentialism, this argument does not strongly depend on whether different cognitive systems evolved via convergent evolution or not.

is, the computational account has implications for the potential explanatory scope of an algorithmic account.

PET is the theory that people's representations of certain substances are structured around the representation of an *essence* (Gelman, 2003; Medin & Ortony, 1989). According to the theory, people (implicitly) represent a feature of the substance, the essence, as being causally responsible for other substance-typical properties, and think that substance membership is mostly determined by possession of the essence. For instance, people think that there is some non-obvious property (the essence of gold) that causes gold to have the other properties that it has (such as being yellow and shiny), and that for something to qualify as gold, it must have that property. People think there is such an essence of gold even when they cannot specify what it is: their representation of "gold" contains an "essence placeholder" (Medin & Ortony, 1989) that exists even before they learn that the essence of gold is atomic number 79.

Note that researchers writing about essentialism are not usually explicit about which level of analysis they are interested in. Nonetheless, within Marr's framework, PET may be most appropriately described as being concerned with the algorithmic level. It is a theory about the structure of the representations over which algorithms for categorization operate; it does not aim at describing this structure in detail, but it does make statements about it. For instance, the theory claims that substance representations are structured around an essence placeholder: this is not a statement about how the mind ought to be (many cognitive scientists actually argue that many substances do not have true essences (Gelman, 2003), which makes it sub-optimal to represent them as such), but a statement about how the mind is structured. Contrary to a computational theory, PET is not concerned with an analysis of the information-processing problems that categorization solves, or with the identification of optimal solutions to these problems. It is not an implementation-level theory either; it does not make statements about, for instance, where essentialism is localized in the brain, or the kind of neural firing patterns that could underlie the phenomenon.[4]

Psychologists have argued that essentialism can explain a wide variety of phenomena, such as: why people attribute more aesthetic value to an original painting than a physically identical copy (Bloom, 2010; Gelman, 2013), why natural selection is counter-intuitive (Shtulman & Schulz, 2008), why people are opposed to GMOs (Blancke, Van Breusegem, De Jaeger, Braeckman, & Van Montagu, 2015), why they believe that every person has a hidden True Self that is morally good (De Freitas, Cikara, Grossmann, & Schlegel, 2017), or why they think that an organ transplant might change someone's personality (Meyer, Gelman, Roberts, & Leslie, 2017). If a single theory about the format of mental representations could explain all essentialist intuitions, then this theory would seem to have a very impressive explanatory scope.

Yet in light of the computational account sketched here, one does not strongly expect that the same kind of representational structure underlies all essentialist intuitions. If essentialist intuitions follow from design requirements of inductive inference systems, then it might be that some of these intuitions are widespread simply because they fall out of widespread functional requirements. Although one might say that people have essentialist intuitions about the Mona Lisa and about tigers, it is still possible that the mental representation of the Mona Lisa and the mental representation of "tigers" have a very different structure. The assumption that one needs a single algorithmic theory to explain essentialism about the Mona Lisa and essentialism about tigers may simply be the result of a neglect of the computational level of explanation.

[4] Some authors (e.g. those rejecting information-processing approaches to psychology) may argue that a theory does not need to belong to any one of Marr levels of explanation (I thank an anonymous reviewer for this remark).

Making this assumption may be the same kind of mistake as that made by a biologist who would assume that the ontogenetic explanation for the existence of bat wings will also apply to butterflies and pterodactyls, because he cannot see the - more general - functional explanation for the existence of wings in these various species. In the absence of a convincing computational account of essentialist intuitions, it may seem that the only explanation for their prevalence is the hypothesis that the cognitive systems underlying these intuitions share some deep structural similarities. By contrast, the proposal that these intuitions follow from basic design requirements of these systems raises the possibility that the systems actually have very different structures. For instance, different cognitive mechanisms may be non-descriptivist *despite* using very different algorithms: the fact that these systems are all non-descriptivist can be explained by the fact that these systems share non-descriptivism as a basic design requirement. The same design requirement can have a variety of alternative solutions at the algorithmic level – therefore it is possible that the same design requirement will be met using different algorithms or different representational formats in different cognitive systems (Marr, 1982).

Note that the computational account does not *eliminate* the need for algorithmic theories: even if one has a computational account of the problem being solved by a given cognitive system, in order to fully understand this system, one also needs to give an algorithmic explanation for how this system produces essentialist intuitions. The present argument is that one may not need to account for all essentialist intuitions with the same algorithmic-level theory. Furthermore, the only explanation that is *general enough* to account for all instances of a class of essentialist intuitions may turn out to be a computational-level explanation.

## 7. A unifying algorithmic theory of essentialist intuitions may be impossible

On functional grounds one expects that different inference systems will in fact have different structures, because specialization is crucial to successful inference (see Section 5 above). In general, different inference systems will have slightly different functional requirements. For instance, a system for the representation of artifact kinds and a system for the representation of biological kinds are expected to make use of different bodies of background knowledge relevant to their respective domains. And indeed the data suggests that they do, since the same task (e.g. the transformation task in Keil, 1989) can elicit different sorts of intuitions in people reasoning about animals vs artifacts.

Therefore, instead of being the symptom of a common representational structure, one expects "essentialism" to be a loose constellation of intuitions. These intuitions arise as the output of systems that solve related, but ultimately distinct, information-processing problems. Given that systems with different design requirements are expected to have different representational structures, one does not really expect these intuitions to be explained by a single algorithmic theory. Some essentialist intuitions may be impossible to explain without appealing to domain-specific knowledge: for instance it is plausible that people's sensitivity to genealogical relationships when reasoning about animals does not just fall out of a general representational structure, but is rather part of a specialized package of knowledge about the biological world. Thus, some essentialist intuitions may require *more* than PET to be fully explained.

Conversely, some essentialist intuitions may be explained as falling out of simpler representational formats. Consider, for example, how the mind represents individual objects and tracks them across time and space. There exists a large literature on the cognitive psychology of object representation, most of which focuses on relatively "mid-level" visual processes, such as those involved when people have to keep track of individual items moving on a screen. In a review of this literature, Leslie, Xu, Tremoulet, and Scholl (1998) conclude that: "the classical idea of object representations as bundles of sensations, perceptual features, or properties of any kind, might be fundamentally mistaken. Instead, the heart of any object representation might be inherently abstract, a kind of mental pointing at a 'this' or at a 'that'." (p.17). This is strikingly similar to the claim in the essentialist literature (Gelman, 2003) that the representation of a natural kind like "tigers" is not a bundle of features, but an abstract pointer to some yet-undiscovered reality. In other words, it seems that non-descriptivism does not only hold for "high-level" processes like categorization on the basis of natural kinds, but also for the "mid-level" processes by which the mind organizes visual experience.

To give a simple example illustrating what Leslie et al. (1998) mean, we think of two forks laying on a table as being two distinct objects, even when we know the forks to be physically identical in every respect, and we would be able to track the identity of each fork even as they are moving. Therefore we do not need to rely on any physical difference between the forks in order to conceive of them as distinct individuals. Convergent lines of evidence have shown that humans spontaneously represent objects in such a non-descriptivist format (Kahneman, Treisman, & Gibbs, 1992; Pylyshyn, 2001, 2007). For instance, when people are asked to track the identity of items moving randomly on a screen, they do so mostly on the basis of spatio-temporal continuity, and pay little attention to persistent perceptual features of the objects: their performance does not decrease when objects change color, size or shape during movement (Pylyshyn, 2001). Non-descriptivism in object representation is already present in infancy (Richardson & Kirkham, 2004; Van de Walle, Carey, & Prevor, 2000; Xu & Carey, 1996).

An explanation of how people represent forks in a non-descriptivist format, according to PET, would be the following: the mind always implicitly attributes a different invisible essence to each fork, and can conceive of them as being distinct only as a result of this operation. However, this may seem like a baroque account of our ability to track objects across space and time. Instead, theories of object representation rely on simpler explanations, such as the fact that object-tracking works on the basis of cues of spatio-temporal continuity (Carey, 2009; Leslie et al., 1998; Pylyshyn, 2007).

Individual forks do not in fact have individual essences, so it is obviously not a design requirement of object representation systems to engage in such ontological commitment; whereas it is a design requirement for this system to be non-descriptivist. By contrast, in order to exploit the complex causal structure of some natural kinds, systems for representing a natural kind category like "gold" are expected to have a more complex structure, possibly coming closer to that posited by PET.

In summary, the existence of a single algorithmic theory that accounts for all essentialist intuitions appears unlikely. PET might very well be a good algorithmic explanation of some subset of essentialist intuitions, but one should be careful not to overestimate its explanatory scope. For some inference problems, the structure described by PET may be more than what is strictly required to solve that problem; for other problems, it might not be enough, or altogether inappropriate.

Attributing to PET more explanatory power than it actually has can be problematic, because illusions of explanatory depth mask the need for deeper understanding of a phenomenon. For instance, the theory that people represent biological kinds as having essences does not predict that people will be *morally opposed* to the act of mixing these essences together, so the theory does not on its own explain anti-GMOs sentiment. Similarly, the theory that people represent individual persons as having essences does not predict that these essences will be represented as morally good, so it is at best an incomplete explanation of "True Self" beliefs. It is likely that only an analysis of the specific information-processing problems involved in the representation of biological kinds, or of persons, can explain why essentialist intuitions about GMOs or about the True Self take the forms that they do.

Note that part of the appeal of a unifying algorithmic theory stems from the fact that essentialism has often been held to explain why people have *strange* beliefs, such as beliefs in the existence of a True Self or in the capacity for organ transplants to change one's personality. It is tempting to think that a normative, computational theory is inconsistent with such phenomena: if people form erroneous beliefs, then their cognitive mechanisms must deviate from the optimal design, and an algorithmic theory like PET is needed to explain how. But this would be a fallacy: one cannot use the fact that a cognitive system makes mistakes to conclude that the system is not optimally designed. As long as an inference system has to go beyond the data given, it has to make informed guesses, and these will produce sound judgments in normal conditions but are bound to produce mistakes in unusual conditions; for instance the visual system has to infer the 3D structure of the world from the 2D image on the retina, and this makes it easy for cunning cognitive scientists to induce the false perception of depth. Similarly, the fact that people's inferences about the deep reality of many entities are often inconsistent with contemporary scientific knowledge does not mean that their minds are flawed. Such mistaken inferences are an unavoidable consequence of the fact that people have to make these inferences on the basis of limited, and sometimes misleading, input.

Note that proponents of PET sometimes speculate that essentialist intuitions are a uniquely human phenomenon (e.g., Gelman, 2003). On the adaptationist account sketched here, one does not expect essentialism to be an idiosyncratic fact about the brain of a single species. Though preliminary, evidence for essentialist intuitions in non-human primates (Cacchione, Hrubesch, Call, & Rakoczy, 2016; Phillips, Shankar, & Santos, 2010) are consistent with the latter account.

## 8. Conclusion

The origins of essentialism in cognition are still seen by many cognitive scientists as "something of a mystery" (Cimpian & Salomon, 2014, p.462). Yet from a computational point of view, they are straightforward: essentialist intuitions are the unsurprising output of well-designed cognitive adaptations for inductive inference.

Variations on this important insight have previously been mostly used to provide a gloss of evolutionary plausibility to theories of cognitive structure, in a post-hoc fashion (e.g., Bloom, 2000; Gelman, 2003). Here I have put forward a sketch of a more adaptationist approach. Because of the nature of the inference problems that many organisms have to solve, one expects them to have cognitive mechanisms that construct representations that are more abstract than simple lists of features, are sensitive to causal relationships, and informed by a rich domain-specific background knowledge. This makes is unlikely that essentialism comes from some special fact about the structure of a restricted set of high-level cognitive processes in the human mind.

## Conflicts of interest

None.

## Acknowledgements

I would like to thank Michael Barlev, Leda Cosmides, Tamsin German, Zoe Liberman, and two anonymous reviewers for helpful suggestions and comments.

## References

Ahn, W. K., Kalish, C., Gelman, S. A., Medin, D. L., Luhmann, C., Atran, S., ... Shafto, P. (2001). Why essences are essential in the psychology of concepts. *Cognition, 82*(1), 59–69.

Ahn, W. K., Kim, N. S., Lassaline, M. E., & Dennis, M. J. (2000). Causal status as a determinant of feature centrality. *Cognitive Psychology, 41*(4), 361–416.

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review, 98*(3), 409.

Atran, S. (1998). Folk biology and the anthropology of science: Cognitive universals and cultural particulars. *Behavioral and Brain Sciences, 21*(04), 547–569.

Barrett, H. C. (2001). On the functional origins of essentialism. *Mind & Society, 2*(1), 1–30.

Blancke, S., Van Breusegem, F., De Jaeger, G., Braeckman, J., & Van Montagu, M. (2015). Fatal attraction: The intuitive appeal of GMO opposition. *Trends in Plant Science, 20*(7), 414–418.
Bloom, P. (1996). Intention, history, and artifact concepts. *Cognition, 60*(1), 1–29.
Bloom, P. (2000). *How children learn the meanings of words.* MIT press.
Bloom, P. (2010). *How pleasure works.* New York: W. W. Norton & Co.
Boyer, P. (1998). If "tracking" is category-specific a "common structure" may be redundant. *Behavioral and Brain Sciences, 21*(01), 67–68.
Boyer, P. (2000). Natural epistemology or evolved metaphysics? Developmental evidence for early-developed, intuitive, category-specific, incomplete, and stubborn metaphysical presumptions. *Philosophical Psychology, 13*(3), 277–297.
Boyer, P., & Barrett, H. C. (2015). Intuitive ontologies and domain specificity. In D. Buss (Ed.). *The handbook of evolutionary psychology* (pp. 161–179). Hoboken, NJ: John Wiley & Sons.
Cacchione, T., Hrubesch, C., Call, J., & Rakoczy, H. (2016). Are apes essentialists? Scope and limits of psychological essentialism in great apes. *Animal Cognition, 19*(5), 921–937.
Carey, S. (1996). Cognitive domains as modes of thought. In D. R. Olson, & N. Torrance (Eds.). *Modes of thought: Explorations in culture and cognition* (pp. 187–215). New York: Cambridge University Press.
Carey, S. (2009). *The Origin of Concepts.* Oxford University Press.
Chomsky, N. (1959). A review of BF Skinner's verbal behavior. *Language, 35*(1), 26–58.
Cimpian, A., & Salomon, E. (2014). The inherence heuristic: An intuitive means of making sense of the world, and a potential precursor to psychological essentialism. *Behavioral and Brain Sciences, 37*(05), 461–480.
Cosmides, L., & Tooby, J. (1994). Beyond intuition and instinct blindness: Toward an evolutionarily rigorous cognitive science. *Cognition, 50*(1), 41–77.
De Freitas, J., Cikara, M., Grossmann, I., & Schlegel, R. (2017). Origins of the belief in good true selves. *Trends in cognitive sciences, 21*(9), 634–636.
Fodor, J. A. (1998). *Concepts: Where cognitive science went wrong.* Oxford University Press.
Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought.* USA: Oxford University Press.
Gelman, S. A. (2013). Artifacts and essentialism. *Review of Philosophy and Psychology, 4*(3), 449–463.
Gelman, S. A., & Coley, J. D. (1990). The importance of knowing a dodo is a bird: Categories and inferences in 2-year-old children. *Developmental Psychology, 26*(5), 796.
Gelman, S. A., & Coley, J. D. (1991). Language and categorization: The acquisition of natural kind terms. In J. P. Byrnes, & S. A. Gelman (Eds.). *Perspectives on language and thought: Interrelations in development* (pp. 146–196). Cambridge: Cambridge University Press.
Gelman, S. A., & Markman, E. M. (1986). Categories and induction in young children. *Cognition, 23*(3), 183–209.
Gelman, S. A., & Markman, E. M. (1987). Young children's inductions from natural kinds: The role of categories and appearances. *Child Development,* 1532–1541.
Gelman, S. A., & Wellman, H. M. (1991). Insides and essences: Early understandings of the non-obvious. *Cognition, 38*(3), 213–244.
Glymour, C. N. (2001). *The mind's arrows: Bayes nets and graphical causal models in psychology.* MIT press.
Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: Causal maps and Bayes nets. *Psychological Review, 111*(1), 3.
Griffiths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive Psychology, 51*(4), 334–384.
Hall, D. G. (1998). Continuity and the persistence of objects: When the whole is greater than the sum of the parts. *Cognitive Psychology, 37*(1), 28–59.
Helmholtz, H. (1925). *Treatise on physiological optics.* Vol. 3; J. Southall, Trans.. Rochester, NY: Optical Society of America (Original work published 1910).
Hirschfeld, L. A. (1995). Do children have a theory of race? *Cognition, 54*(2), 209–252.
Inagaki, K., & Hatano, G. (2006). Young children's conception of the biological world. *Current Directions in Psychological Science, 15*(4), 177–181.
Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology, 24*(2), 175–219.
Keil, F. C. (1989). *Concepts, kinds, and cognitive development.* Cambridge: MIT Press.
Kemp, C., & Tenenbaum, J. B. (2008). The discovery of structural form. *Proceedings of the National Academy of Sciences, 105*(31), 10687–10692.
Knill, D. C., & Richards, W. (Eds.). (1996). *Perception as Bayesian inference.* Cambridge University Press.
Leslie, A. M., Xu, F., Tremoulet, P. D., & Scholl, B. J. (1998). Indexing and the object concept: Developing what and where systems. *Trends in Cognitive Sciences, 2*(1), 10–18.
Markman, E. M. (1989). *Categorization and naming in children: Problems of induction.* MIT Press.
Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information.* San Fransisco, CA: Henry Holt & Company.
Medin, D. L., & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou, & A. Ortony (Eds.). *Similarity and analogical reasoning* (pp. 179–195). Cambridge, England: Cambridge University Press.
Meyer, M., Gelman, S. A., Roberts, S. O., & Leslie, S. J. (2017). My heart made me do it: Children's essentialist beliefs about heart transplants. *Cognitive Science, 41*(6), 1694–1712.
Millikan, R. G. (1998). A common structure for concepts of individuals, stuffs, and real kinds: More mama, more milk, and more mouse. *Behavioral and Brain Sciences, 21*(01), 55–65.
Newman, G. E., Herrmann, P., Wynn, K., & Keil, F. C. (2008). Biases towards internal features in infants' reasoning about objects. *Cognition, 107*(2), 420–432.
Newman, G. E., & Knobe, J. (2018). The essence of essentialism. *Mind & Language* (in press).
Pearl, J. (2000). *Causality: models, reasoning and inference.* Cambridge University Press.
Phillips, W., Shankar, M., & Santos, L. R. (2010). Essentialism in the absence of language? Evidence from rhesus monkeys (*Macaca mulatta*). *Developmental Science, 13*(4), F1–F7.
Pylyshyn, Z. W. (2001). Visual indexes, preconceptual objects, and situated vision. *Cognition, 80*(1), 127–158.
Pylyshyn, Z. W. (2007). *Things and places: How the mind connects with the world.* MIT press.
Recanati, F. (2012). *Mental files.* Oxford University Press.
Rehder, B. (2007). Essentialism as a generative theory of classification. In A. Gopnik, & L. Schultz (Eds.). *Causal learning: Psychology, philosophy, and computation* (pp. 190–207). Oxford, England: Oxford University Press.
Richardson, D. C., & Kirkham, N. Z. (2004). Multimodal events and moving locations: Eye movements of adults and 6-month-olds reveal dynamic spatial indexing. *Journal of Experimental Psychology: General, 133*(1), 46.
Rips, L. J. (1989). Similarity, typicality, and categorization. In S. Vosniadou, & A. Ortony (Eds.). *Similarity and analogical reasoning* (pp. 21–59). New York: Cambridge University Press.
Schulz, L. (2012). The origins of inquiry: Inductive inference and exploration in early childhood. *Trends in Cognitive Sciences, 16*(7), 382–389.
Setoh, P., Wu, D., Baillargeon, R., & Gelman, R. (2013). Young infants have biological expectations about animals. *Proceedings of the National Academy of Sciences, 110*(40), 15937–15942.
Shtulman, A., & Schulz, L. (2008). The relation between essentialist beliefs and evolutionary reasoning. *Cognitive Science, 32*(6), 1049–1062.
Smith, E. E., & Medin, D. L. (1981). *Categories and concepts.* Cambridge, MA: Harvard University Press.
Strevens, M. (2001). Only causation matters: Reply to Ahn et al. *Cognition, 82*(1), 71–76.
Tenenbaum, J. B., Griffiths, T. L., & Niyogi, S. (2007). Intuitive theories as grammars for causal inference. In A. Gopnik, & L. Schultz (Eds.). *Causal learning: Psychology, philosophy, and computation* (pp. 301–322). Oxford, England: Oxford University Press.
Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science, 331*(6022), 1279–1285.
Tinbergen, N. (1963). On aims and methods of ethology. *Ethology, 20*(4), 410–433.
Tooby, J., & Cosmides, L. (1992). The psychological foundations of culture. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.). *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 19–136). New York: Oxford University Press.
Van de Walle, G. A., Carey, S., & Prevor, M. (2000). Bases for object individuation in infancy: Evidence from manual search. *Journal of Cognition and Development, 1*(3), 249–280.
Welder, A. N., & Graham, S. A. (2001). The influence of shape similarity and shared labels on infants' inductive inferences about nonobvious object properties. *Child Development, 72*(6), 1653–1673.
Xu, F., & Carey, S. (1996). Infants' metaphysics: The case of numerical identity. *Cognitive Psychology, 30*(2), 111–153.