Refining and expanding the proposal of an inheritance heuristic in human understanding

doi:10.1017/S0140525X14000028

Andrei Cimpian and Erika Salomon
Department of Psychology, University of Illinois, Champaign, IL 61820.
acimpian@psych.illinois.edu salomon3@illinois.edu
http://psychology.illinois.edu/people/acimpian
http://www.erikasalomon.com

Abstract: The inheritance heuristic is a cognitive process that supplies quick and effortless explanations for a wide variety of observations. Due in part to biases in memory retrieval, this heuristic tends to overproduce explanations that appeal to the inherent features of the entities in the observations being explained (hence the heuristic’s name). In this response, we use the commentators’ input to clarify, refine, and expand the inheritance heuristic model. The end result is a piece that complements the target article, amplifying its theoretical contribution.

We have used the commentators’ questions to expand the original statement of our inheritance heuristic (IH) model. Thus, we intend this response to serve as a substantive companion piece to our target article, with signposts along the way to indicate which commentaries prompted which clarification, refinement, or expansion. We are truly grateful for everyone’s input—the diversity of perspectives reflected in the commentaries has given our theory a richness that would have been unattainable otherwise.

The organization of this response loosely mirrors that of the target article, in that we start with a discussion of the IH process and then move on to the phenomena (including essentialism) that we have claimed to stem from this process.

R1. The inheritance heuristic

This section capitalizes on the commentators’ questions to elaborate our account of the process that underlies the IH.

R1.1. The inheritance heuristic as a general explanatory heuristic process

Most broadly (and boldly) stated, the goal of our account is to outline the general process by which many everyday explanatory intuitions are generated. A key claim of the account is that the process of generating such intuitions has much in common with the heuristic judgments people make in response to other difficult questions (e.g., Gilovich et al. 2002; Kahneman 2011). Specifically, this process overuses information that is easily accessible, ignoring other relevant considerations. Although questions about why the world is a certain way (e.g., why do girls like pink?) are incredibly complex (the sorts of questions that scientists spend entire careers researching), plausibility-seeming answers come to mind with surprising ease. How is it that people are not stumped? How do they come up with any sort of answer (let alone an almost instantaneous one), given the daunting complexity of the explanatory task? Our theory offers a solution to this puzzle: We propose that people rely on a heuristic shortcut that leads them to explain by using only the information that’s most accessible to them. Because this information often consists of the inherent features of the entities under consideration, the explanations generated will be correspondingly skewed toward inheritance—hence the term inheritance heuristic.

Far from being just another quirk of human cognition, this heuristic is likely to have a powerful influence on people’s understanding of the world. People seem motivated to wonder why from the youngest ages (e.g., Anderson et al. 1996; Callanan & Oakes 1992; Gopnik et al. 2004; Murphy & Medin 1985; Schulz 2012); thus, if our intuitive answers are shaped from the youngest ages by the proposed inheritance heuristic, then the possibility of pervasive bias is high.

To reiterate, the sort of heuristic we are proposing is a general explanatory process that avails itself of the most readily accessible knowledge to generate its output and that, as a result, ends up overusing inherent features. Thus, the IH is not “devoted to” postulating inherent explanations (Dunham, para. 4); it inadvertently (over)postulates these explanations because inherent features are so accessible in so many circumstances. But there is nothing in the structure of the IH that prevents it from using salient, easily accessible extrinsic information (e.g., about ownership; see Noles & Danovitch) to generate an
explanation. The output may not be an inheritance-based explanation in this case, but that is not inconsistent with our account: The proposed heuristic is an inheritance heuristic not because it is somehow structurally incapable of handling anything other than inheritance, but simply because it tends not to, given that other sorts of information aren’t typically at its “fingertips” (to use Dunham’s evocative metaphor [para. 2]). In effect, the IH and the explanatory heuristic described by Dunham in his thoughtful commentary are one and the same. We thank him for motivating this clearer, and more forceful, statement regarding the nature of the IH process.

This discussion has implications for how the label inheritance heuristic should be understood. The inclusion of inherence in this label is intended simply as a reminder of the typical output of the heuristic process we are proposing: inherence is not the goal of this process, nor is it a criterion that’s explicitly built into its structure. In other words, the term inheritance heuristic describes the entire process illustrated in Figure 2 of the target article, whatever the output of the process may be. This is not just splitting hairs: Misunderstanding the name of the heuristic can lead to misconceptions about what we are proposing. For example, Khalidi & Mugg seem to have interpreted the name of the heuristic as signaling that the IH somehow selects inherent features for inclusion in explanations because of their inherence. This interpretation led Khalidi & Mugg to disagree with our claims, even though we suspect there is no real disagreement between us. Specifically, they questioned the existence of an inheritance heuristic while simultaneously endorsing our main claim: that inherent features are used more often than extrinsic ones in explanations because the former are more accessible. Thus, their skepticism is directed toward a heuristic we didn’t propose—one that selects explanatory features for their inherence per se.

As an aside, all of this reminds us of Kahneman and Frederick’s (2002) remark that “the label that is chosen for a heuristic may take on a life of its own in subsequent theorizing” (p. 65). The process we are proposing is most transparently and accurately labeled with something along the lines of the explanatory heuristic process that tends to output explanations that appeal to inherent facts because these facts are often the most accessible. For obvious practical reasons, we need something a little more succinct, and the inheritance heuristic seems to fit the bill. Although we did not provide sufficient justification for this term the first time around, we hope to have gotten ahead of the curve now and prevented further confusion about its intended meaning.

In response to points raised by the commentators, we go on to highlight several key aspects of the IH process. As a general explanatory heuristic, the IH is triggered by a wide range of explananda. Thus, the answer to Kinzler & Sullivan’s question about whether the IH could generate explanations both for specific instances/events and for broader uniformities is affirmative. The IH is triggered by both and likely leads to intuitions biased toward inherence in both cases. The extent of the bias, however, may be greater for broad patterns than for specific instances because of differences in what information is at its fingertips in these two cases. Extrinsic information is more readily available when the retrieval cue is a specific object or individual: Think of ownership, for example, or of causal history. Such information is more naturally represented and stored at the level of instances, not broad sets or categories. It’s Fido that belongs to our family, that caught a squirrel yesterday, or that was adopted from the shelter—not dogs as a category. The more extrinsic information accessible to the shotgun, the more frequently this information will end up in the explanatory intuitions generated, leading in the case of particular instances to an attenuation (but most likely not elimination) of the usual skew toward inherence relative to what is normatively warranted. This attenuation should not be taken to signal that the IH is solely, or even preferentially, invoked for patterns (as Gelman & Meyer and Prasada state). On our account, the IH process is invoked, and produces explanations, for all sorts of explananda, both general and specific; what differs between these cases is simply the relative extent to which inherent features dominate its output.

While on the topic of the IH as a heuristic process, we agree with Wood that much of the cognitive underbelly of the IH is inaccessible to conscious introspection. The IH is an intuitive (or System 1) heuristic, and lack of conscious access to its workings is a typical feature of such a process (for a nuanced discussion, see Evans & Stanovich 2013). However, we don’t agree with Wood’s stronger claim that the IH operates as a Fodorian peripheral module (Fodor 1983). Although it may be similar to a module in its opaqueness to introspection, the IH does not display many of the other key characteristics of such a module. For instance, there is no reasonable sense in which the IH is domain specific; rather, it operates on an incredibly broad range of inputs. The IH is not informationally encapsulated either; in fact, because it has free access to the information in semantic and episodic memory, the IH cannot be farther from encapsulation.

Although we are denying that the IH itself is a domain-specific explanatory module, we acknowledge that such modules may exist and may provide ready-made explanations for select chunks of our experiences with the world. For example, humans’ ability to make sense of the interactions between physical objects seems to be facilitated by a trove of early-developing perhaps even innate, causal-explanatory principles (e.g., Baillargeon 2004; Spelke & Kinzler 2007). More recent evidence suggests that a similar explanatory module may be available for making sense of the interactions between psychological agents (e.g., Kovác et al. 2010; Onishi & Baillargeon 2005). Thus, our claim that the IH can account for many everyday explanatory intuitions should not be misunderstood as stronger than it is. The IH is clearly not the only explanatory game in town, and we suspect that its influence will be felt most strongly in the (many) circumstances where these domain-specific explanatory principles are silent.

R1.2. What’s a heuristic, anyway?

Because there was some disagreement in commentators’ assumptions about what a heuristic is (Bookstein; Braisby; Dubljević & Racine; Gaucher & Jost), we should clarify where we stand on this issue. Following a widely accepted view in the recent literature on heuristics and biases, we understand a heuristic to have at its core a form of question substitution (e.g., Kahneman 2011; Kahneman & Frederick 2002). When trying to come up with solutions to complex problems or questions, people
inadvertently rely on easily accessible information and/or simple computations more often than would be normatively warranted. This liberal reliance on cognitive “low-hanging fruit” enables people to find quick answers to questions that are enormously complicated, which is arguably a better outcome than just being perplexed. But there’s a catch: The answers generated via this process no longer address the original questions but related, and cognitively simpler, ones. Thus, from an observer’s perspective, it seems as though people often substitute easy questions for harder ones. For instance, when asked a question about the probability that Tom W. is a computer scientist, people often answer as though they had substituted a related, easier question about Tom W.’s similarity to the stereotypical computer scientist (Kahneman & Tversky 1973). To clarify, though, question substitution is just a descriptive tool – another metaphor. No question-swapping is actually (i.e., causally) involved in the process by which people generate their heuristic answers; the key feature of this process is its reliance on easily accessible information. However, this reliance leads to responses that make it seem as though people are answering a different question than the one they set out to answer, which justifies the metaphor of question substitution.

Responses that can be characterized as question substitutions have been identified in a wide array of complex judgments: People often judge stimulus familiarity as though they were rating stimulus positivity/attraction instead (e.g., Monin 2003); they seem to decide on appropriate punishments as though they were simply reporting how much outrage they felt at the crimes in question (e.g., Sunstein 2005); and so on. The IH model proposes that our everyday explanatory judgments often involve a similar question substitution: The complex question “What explains this observation?” is often answered (due to the high accessibility of inherent facts) as though it were the easier question “What inherent facts explain this observation?”

A corollary of this question-substitution view of the IH is that its output is – at best – an approximation of the truth. Thus, we are unmoved by examples where the IH seems to lead to mistaken judgments. (Both Braisby and Dubljević & Racine bring up such examples.) These would count as counterevidence to our claims only if we assumed that the IH is perfectly accurate. Since many of our own examples (in the target article) point to the suspect intuitions generated by the heuristic, we hope it is clear we do not endorse such an assumption. However, we are agnostic about exactly how often the heuristic will output the correct answer (that is, beyond the weak claim that it isn’t always right). Note that we can afford to remain agnostic on this point because the truth of our hypothesis is independent of the frequency with which the IH is right: Our proposal is of a specific cognitive process by which everyday explanations are generated – a process that can be described as a sort of question substitution. This mechanistic proposal should be evaluated on its own merits, without getting mired in ideological debates about heuristic accuracy (e.g., Gigerenzer 1996; Kahneman & Tversky 1996).

**R1.3. What exactly are inherent features?**

Our account makes a basic distinction between inherent and extrinsic facts. To elaborate on the definition provided in our target article, inherent facts are those that can be said to characterize a thing “in virtue of the way that thing itself, and nothing else, is” (Lewis 1983, p. 197). The shape of an orange is an inherent fact about it; the fact that it is found next to some apples is not, and neither is the fact that it was washed in a sink – these are extrinsic facts about it. Another intuitive way to understand this distinction is that inherent facts about an object are those facts that must be true about a perfect duplicate of that object. A perfect duplicate of an orange would necessarily have the same shape as the original (as well as the same color, DNA, chemical composition, etc.); it would not, however, have the same location or history.

In part because of our cursory description of this distinction in the target article, several of our commentators understood us to mean something else by inherent; almost all of these alternative meanings for the term were more restrictive than our intended meaning. Strevens thought we meant inherent to be synonymous with internal. (Note, however, that a feature such as the shape of an orange is inherent without being internal.) Similarly, Noles & Danovitch thought that inherent properties are internal and nonobvious, whereas Hampton thought they are necessarily “deep.” Prasada defined as inherent only those features that have a principled or causal connection with their kinds (e.g., Prasada et al. 2013). At the other end of the spectrum, Khalidi & Mugg understood precisely the distinction we were trying to make but then went on to use inherent as if it were tantamount to perceptually salient. (However, a feature such as the chemical composition of an orange is inherent without being perceptually salient.) In light of the foregoing expanded definition, it should be clear that none of these alternative definitions adequately capture the distinction we are making.

We should also clarify that, as Khalidi & Mugg helpfully pointed out, our account operates with a psychological notion of inherent rather than a metaphysical one. That is, our hypothesis does not assume that inherent properties actually exist – that they are real aspects of the world (a metaphysical claim). This is just as well, since the metaphysical notion of inherent has proven difficult to pin down (e.g., Weatherson & Marshall 2013): Many properties might at first appear inherent, but it is in fact unclear whether they are really (that is, metaphysically) so. For example, is an object’s shape an inherent feature of that object, or is it an extrinsic feature because it depends on the curvature of the space around the object? However, our account can sidestep these metaphysical difficulties. Intuitively, the shape of an object appears to most people to be an inherent feature of that object, and that is all we need here.

Thus, our account assumes only that people intuitively conceive of some features as inherent and of others as extrinsic (a psychological claim). Note, however, that even this psychological assumption is needed in only a limited sense: The IH relies on inherent features not because of their inherent per se but rather because they tend to be low-hanging fruit for retrieval processes. Therefore, if inherent features are used just because they come to mind more readily than other facts, then the heuristic process doesn’t actually need to be able to pick out inherent features. In other words, people’s intuitive ability to distinguish between inherent and extrinsic facts is not directly involved in the process by which the heuristic generates its output.
So, then, what work does the inherent/extrinsic distinction do in our account? Most obviously, it is a descriptive tool that enables us (as researchers) to characterize the output of the hypothesized explanatory heuristic and to determine whether, as claimed, it typically relies on inherent facts. This distinction may also play a mechanistic role, but only after the IH has generated its output. Specifically, the ability to identify inherent features as such may enable reasoners to take the output of the IH and draw additional inferences from it, depending on whether it appeals to inherent or extrinsic facts. For instance, crucial inferences about the stability, inalterability, and naturalness of the phenomena explored are considerably more likely to follow from explanations that (are understood to) appeal to inherent facts.

But can people in fact draw the inherent/extrinsic distinction consistently? Is there sufficient agreement as to what features are inherent and which are extrinsic? Although this question awaits further investigation, the evidence so far leads us to believe that the variability in people’s intuitions about inherence is relatively minimal, contrary to Khalidi & Mugg’s claim. Khalidi & Mugg raise the possibility of widespread disagreement about these matters via several examples. For instance, do people view features such as orange juice’s being tangy or healthy as inherent (as we claimed), or do they view them as extrinsic because they depend on a relation between OJ and humans? The empirical data we know of point to the conclusion that these sorts of features are in fact consistently judged to be inherent: People think that lemons are sour, that carrots are crunchy, that snow is white, and so on, simply by virtue of being the kinds of things they are (Prasada & Dillingham 2006) and not because of the relations they bear to humans (specifically, to our perceptual apparatus). We don’t doubt that ambiguous cases exist, but they should not detract from the conclusion that making intuitive inherent/extrinsic judgments is typically straightforward.

**R1.4. Is the inherence heuristic an inherent feature of the human mind?**

Some of our commentators raised the question of whether the IH is itself an inherent feature of cognition (Baron; Bookstein; Hampton; Khalidi & Mugg; O’Connor & Joffe). As far as we are concerned, the answer is “yes and no.” Some elements of the IH process are supplied by the endogenous structure of the human mind, whereas others are filled in by the external world. Among the likely candidates for the inherent components of the IH are the human propensity to ask *why*, the fallibility of our memory (which makes it so that stored information is not uniformly accessible), and our bounded rationality (which, among other things, leads to a satisfying reliance on easily accessible information) (e.g., Simon 1982). Nevertheless, the operation of the IH also depends in crucial ways on factors extrinsic to the mind. For instance, what information is most accessible to the IH process is in part a function of the multilayered context in which people are embedded (e.g., broad cultural beliefs, specific prior experiences; see sects. R1.5 and R1.6). How people handle the output of the IH (adopt vs. question/revise), and thus the heuristic’s ultimate impact on their belief systems, may likewise be influenced by the sociocultural context (see sect. R1.7).

In sum, as with most other psychological processes, the functioning of the IH depends both on features of the mind and on features of the world.

**R1.5. Which aspects of the world are noticed and encoded? Which are explained?**

A number of commentators (Braisby, Kinzler & Sullivan; Rhodes) raised questions about the steps that precede the IH process per se: What aspects of the world do people notice and encode? And, of these, which do they seek to explain? These questions are clearly important for a comprehensive account of how the IH shapes human understanding across different domains of experience. Nevertheless, full discussion of these matters goes far beyond the scope of this response and, indeed, beyond the limits of current scientific knowledge. Here, we provide only a few speculative thoughts.

Regarding what people attend to, we suspect that the final answer will point to at least two sources of influence. First, human attention is guided by skeletal, domain-specific biases, arguably the products of natural selection, that highlight certain aspects of experience over others (e.g., Baillargeon et al. 2010; Chomsky 1959; Gelman 1990; Izard et al. 2009; Morton & Johnson 1991). Second, the social environment provides another powerful guide for attention, leading to preferential encoding of some observations over others, equally available, ones. In fact, several of our commentators have done groundbreaking work on exactly this topic (e.g., Bigler & Liben 2006; 2007; Gelman et al. 2010; Rhodes et al. 2012). Such sociocultural cues combine with the skeletal biases just mentioned to pare down the vast amount of information that is in principle available to humans. The end product of this filtering process is the information actually encoded, which is the raw material for the IH—what the heuristic is invoked to explain and what it explains with.

Importantly, however, not every bit of information that is encoded triggers the search for an explanation; the fact that humans are motivated to explain and understand does not entail that they will try to explain everything. Undoubtedly, some information is stored in semantic memory without an explanation. (Unlike Braisby, we don’t see this claim as particularly controversial or problematic for our broader account.) The question then becomes, what does trigger the search for an explanation? Which facts do people seek to explain? Unfortunately, this issue has received less empirical attention than the one regarding what information people tend to encode. Some evidence suggests that children are particularly motivated to explain broad facts about the world (Cimpian & Petro 2014), whereas the findings in other studies have suggested that unexpected observations also prompt more explanation seeking (e.g., Legare et al. 2010; Weiner 1985). This cannot be the whole story, though, so much remains to be investigated here.

To conclude, the explanatory heuristic process that we are proposing is triggered by *why* questions about aspects of the world that people have, at some point, noticed. Thus, our proposal will ultimately need to dovetail with research on (a) what prompts people to ask *why*, and (b) what things people tend to notice, in order to comprehensively articulate how the IH affects human understanding across a wide variety of contexts and domains.
**R1.6. What’s in the shotgun?**

The first stage in our model is the mental shotgun: the fast, shallow memory search for relevant facts performed as soon as the IH process is set into motion. As with intuitive heuristics more generally, the operating principle of this stage is to take the path of least cognitive resistance. Here, this operating principle leads the IH to retrieve only the subset of relevant information that is most easily activated from memory. What sort of information might this be? Are there any commonly used in the content of this most-accessible information that hold across the observations being explained, the situations in which the explanation is being generated, and so on? A key insight of the IH model is that there may indeed be some systematicity in the content retrieved by the shotgun. Specifically, it is possible that—on average—the most easily activated information will consist of inherent facts about the entities in the observation being explained (sect. 2.2 of the target article for the full argument). This imbalance in retrieval then propagates through the other stages of heuristic processing, ultimately causing an inherence skew in the intuitive explanations that the IH outputs. Notice the qualifier, though: on average. The shotgun cares only about accessibility, so if extrinsic facts happen to be most accessible for a particular observation or in a particular context, then they will be retrieved just as readily as inherent facts.

Armed with this description of the shotgun, we go on to address the questions that were raised about this stage of the IH. First, in response to comments by Dunham and by Gelman & Meyer, we reiterate that the shotgun does not select facts for their inherent. Rather, it selects facts for their accessibility, which leads only indirectly to the retrieval of a preponderance of inherent facts. Thus, the shotgun is perfectly capable of retrieving (and often does retrieve) extrinsic information; it does not need to be “limited to only inherent features” (Noles & Danovitch, para. 5). Second, the content retrieved by the shotgun will undoubtedly be influenced by the local and broader (e.g., cultural) context in which the IH is deployed. If extrinsic forces (e.g., situations, upbringing) figure prominently in the discourse of one’s cultural community, for example, then they will be more accessible to the shotgun. The same goes for many inherent factors, too, especially less concrete ones. Self-esteem, giftedness, bodily humors, etc., are (or were) all culturally prominent inherent features that are easily retrievable by people belonging to the relevant communities in their search for explanations. In effect, the content retrieved by the shotgun will be influenced by anything that influences the accessibility of explanation-relevant information in memory, either temporarily or chronically. As such, cultural input is certain to influence the IH process. We thank several of our commentators (Baron; Kinzler & Sullivan; O’Connor & Joffe; Olivola & Machery; Yzerbyt & Demoulin) for prompting us to elaborate on this important point.

**R1.7. How does the storyteller work?**

The next stage in the heuristic process is the storyteller: the search for a quick way to assemble (at least some of) the facts pulled up by the shotgun into a plausible explanatory story. The operating principle of the storyteller is the same as that of the shotgun—namely, take the path of least resistance. Thus, the storyteller is likely to capitalize on whatever explanatory framework comes to mind most readily that can organize the information pulled up in the preceding shotgun search. This stage is also likely to terminate its search for an explanation as soon as a first plausible explanation is assembled.

A simple way to think about the relation between the shotgun and the storyteller is that the shotgun supplies the content of the explanations generated by the IH, whereas the storyteller supplies the structure of these explanations. This point brings to light the relative centrality of these two stages to the IH proposal. Because we are fundamentally making an argument about the typical content of everyday explanations, it is the shotgun stage that’s the linchpin of this proposal. The shotgun is where the bias toward inherence originates before propagating through the rest of the IH process: If the content retrieved by the shotgun is inherence-skewed, the storyteller’s options will be constrained accordingly. That is, the content retrieved will be more compatible with, and will facilitate activation of, certain explanatory structures over others. For instance, if the shotgun retrieves a number of inherent features (e.g., the refreshing smell of OJ or its tartness), the storyteller will be hard-pressed to formulate an extrinsic–historical explanation out of such features and may instead appeal to, say, their causal powers. The constraining influence of content on structure is not deterministic, but it may nevertheless be strong enough for inheritance in the shotgun to translate, more often than not, into inherence at the level of the final explanations generated by the IH. (An interesting possibility, which we will mention only briefly here, is that the storyteller might be more than a neutral participant in the IH process—that this stage might actually add to the skew toward inherence rather than simply propagating it. For example, it is possible that inheritance-based explanations are cognitively simpler to build than extrinsic ones, which might often require putting together more elements [e.g., the multiple links in the chain that connects a historical event with a current societal pattern]. If so, then the storyteller’s emphasis on finding a quick solution might lead it to favor the easy-to-assemble inherent explanations, which might in turn amplify the heuristic’s tendency to output such explanations.)

Because the storyteller is somewhat less central to the thrust of our proposal, we were less specific in the target article about its functioning, except to emphasize that the value it places on speed makes it relatively promiscuous in the explanations it generates and deems plausible. Importantly, however, additional details about the operation of the storyteller can be borrowed from the research exploring the structural aspects of explanations (for reviews, see Anderson et al. 1996; Keil 2006; Lombrozo 2006; 2012; Ross 1977). Whatever this research will ultimately conclude about how explanations are typically structured, about the factors that affect their plausibility, and so on, will probably apply to the functioning of the storyteller as well.

Here, we augment our description of the storyteller with a few results from this literature in order to address specific questions brought up by our commentators. First, Marmodoro, Murphy, & Baker (Marmodoro et al.) wondered where the storyteller’s stories come from; seemingly, our account requires that “the causal story preexist in the mind of the reasoner” (para. 2). For example, why
would the storyteller focus on the tanginess of OJ as an explanation for morning consumption if it didn’t already know at some level that tanginess might cause waking? Marmodoro et al. are right in assuming that the storyteller uses previous beliefs in assembling its explanations. However, these beliefs are often quite abstract – for example, sometimes they take the form of general mappings between certain types of causes and certain types of effects. Most relevant to the OJ case, one such abstract mapping may be that sharp perceptual stimuli (e.g., loud noises, bright lights) cause physiological arousal (e.g., being startled, being woken). By subsuming a particular explanandum (e.g., OJ for breakfast) under a more general pattern such as this one, the storyteller can generate an explanation without having a prestored answer. People view the act of subsuming an observation under a general pattern to be explanatory (e.g., Lombrozo 2006; 2012), so there is no reason the storyteller should not be able to avail itself of this strategy in formulating its output.

Second, Prasada provided a plausible example of how the storyteller’s operation might be influenced by the content retrieved by the shotgun. Specifically, Prasada suggested that the storyteller’s output might vary depending on the representation of the inherent features at its disposal. For instance, inherent features that are represented as having a principled connection with their kinds (e.g., tanginess is an aspect of what it means for something to be OJ) might be more likely to license explanations with a more normative flavor (e.g., it’s ideal/right that OJ is for breakfast), as well as downstream intuitions about inevitability. We endorse Prasada’s prediction, and more generally we see great value in research that would further specify how the content pulled up by the shotgun shapes the storyteller’s output.

Third, Rakoczy & Cacchione state that the storyteller’s “complex inferential machinery” is unlikely to be available early in development, in part because “it appears to rely heavily on linguistic capacities” (para. 1). Although we agree that the storyteller is indeed a complex piece of inferential equipment, the jury is still out on whether it (or something like it) is available early in development. Much of the developmental evidence in our target article (e.g., Cimpian & Markman 2009; 2011) relied on language only because it is relatively easy to assess children’s explanations verbally. The obvious drawback of this methodology, though, is that it prevents investigation of these issues in children who haven’t yet acquired language. Nevertheless, the absence of evidence in younger children should not be taken as evidence that the relevant abilities are absent. There is nothing in the structure of storyteller, or the IH process more generally, that depends on linguistic abilities, so we predict that this heuristic process might influence reasoning even before these abilities are in place. In making this prediction, we are also encouraged by the evidence that infants’ general-purpose explanatory abilities are quite sophisticated (e.g., Baillargeon 1994). Although this evidence does not speak directly to our claims, it is at least consistent with the early presence of the sort of inferential machinery that the IH relies on.

Finally, we should reiterate that, just as the content retrieved by the shotgun is influenced by context, so is the operation of the storyteller. For instance, to the extent that different cultures privilege different explanatory frameworks (e.g., Keil 2006), the differential accessibility of these frameworks would likely be reflected in the output of the storyteller. Whether this context-driven variability in the operation of the IH process is sufficient to account for the full range of cross-cultural variability in explanatory practices remains to be established. Unlike Baron, however, we see no need to appeal to multiple separate explanatory mechanisms to account for such variability if a single, context-sensitive mechanism can do the job just as well. (This issue is separate from that concerning whether there exist separate innate, domain-specific explanatory mechanisms, whose important role we have already acknowledged. These mechanisms are too limited in application to account for the range of observations that fall under the scope of the IH, and thus we assume that they are not the explanatory mechanisms Baron proposes as alternatives to the IH.)

R1.8. How can the typical output of the inheritance heuristic be blocked or revised?

Given the prejudicial consequences of the typical (inheritance-based) output of the IH, especially when applied to social groups, a number of commentators have emphasized the importance of understanding how to block or revise this output (Bigler & Clark; Kinzler & Sullivan; Mealey & Crisp; Rhodes). Before discussing their specific suggestions, we provide some preliminary remarks. The concern here is with overriding what we have claimed is the typical output of the IH—namely, its inheritance-based output. Although the IH (as an explanatory heuristic process) can output extrinsic explanations as well, these are not the focus of this section. A reasoner can override the inheritance-based output of the IH by blocking it or by revising it. The difference between these two terms maps onto a difference in the stage at which inheritance is defeated: Blocking occurs when inheritance is avoided before the IH has run its course. Blocking might occur, for example, in circumstances where extrinsic facts happen to be easily accessible and are thus retrieved by the shotgun. Blocking is not effortful (contra Dunham) but rather as effortless-seeming as the rest of the IH process—in fact, blocking occurs during the IH process. In contrast, revisions occur when inheritance-based explanations are overturned after they are generated, and thus after the IH process has run its course. As Rhodes points out, revising already-generated explanations can indeed be effortful, as well as difficult to achieve successfully. We therefore second her call to focus empirical efforts on means of blocking the inheritance-based explanatory “stories” (particularly with respect to social groups, where they are wrong quite often) before they are even told.

In what follows, we briefly highlight three promising suggestions for how to block or revise the inheritance-based output of the IH that are provided by three different sets of commentators. First, Bigler & Clark suggest that perhaps the most effective means of overriding inheritance is also the most direct: providing children with the relevant extrinsic facts, especially in ways that make these facts salient and accessible (e.g., via interactive learning). Consistent with the effectiveness of this inheritance-reduction strategy, Bigler and Wright (2014) review a number of studies in which teaching children about discrimination and prejudice (as extrinsic societal forces) led to corresponding reductions in children’s tendency to attribute
societal phenomena to inherent causes. Interventions such as these may help both by leading children to revise existing inherence-based beliefs and by blocking further such beliefs from being generated.

Second, Kinzler & Sullivan raise the possibility that diverse cultural and linguistic experiences may likewise rely on inherence-based intuitions. Kinzler & Sullivan’s proposal seems to work on two levels: Diverse experiences may increase the diversity of the information accessible to the shotgun, increasing the probability that inherence-based explanations are blocked before they are generated. In addition, having to process varied experiences may, over time, also enhance processing resources such as cognitive control (e.g., Bialystok et al. 2004; Kovács & Mehler 2009), which might increase children’s ability to scrutinize the output of the IH rather than simply accepting it as is. In turn, such scrutiny may increase the likelihood of revisions to inherence-based intuitions.

Third, similarly to Kinzler & Sullivan, Mealey & Crisp suggest that chronic exposure to violations of one’s inherence-based notions (e.g., in diverse environments where reasoners can interact positively with members of outgroups) may predispose one to question the value of these heuristic intuitions and, as a consequence, adopt a more analytic cognitive style. This style may also be facilitated by improvements in cognitive resources brought about by having to repeatedly resolve the conflict between one’s prior beliefs and the evidence contradicting them.

As Mealey & Crisp rightly pointed out, our original statement of the IH model left unspecified the mechanisms by which the typical inherence-based output of this heuristic could be overcome. The foregoing suggestions expand this dimension of the account, laying the groundwork for new empirical research and for interventions to curb the unwanted side effects of inherence-based thinking.

R1.9. Where do inherence-based explanations come from?

Two commentaries question the psychological reality of the heuristic process we have proposed, arguing that inherence-based explanations emerge via other processes. Gelman & Meyer suggest that these explanations stem from a realist assumption about patterns in the world, whereas for Yzerbyt & Demoulin their source can be found in a basic need for coherence and predictability. We will address each of these possibilities in turn but begin with a few general thoughts about the psychological reality of the IH process.

For us, one of the most appealing features of the IH proposal is its non-mysterious nature – its ability to account for an impressive range of cognitive behavior with a simple process whose components are all well understood and independently documented and whose psychological reality is beyond doubt. In a nutshell, we are proposing that why questions trigger a search for relevant information in memory, that this memory search pulls up the information that is most easily accessible, which is typically skewed toward inherent rather than extrinsic (e.g., historical, contextual) facts; and that, due to a tendency to satisfy and to lax supervision by working-memory-dependent (System 2) processes, this retrieval bias toward inherent ends up translating into a bias in the final product of the cognitive system (i.e., explanations). The empirical foundation for each of these components is solid enough that it actually seems difficult to imagine how the process of generating explanations could proceed without them: It seems difficult to imagine how the search for explanations could proceed without a memory search, how this memory search could proceed without sensitivity to well-documented asymmetries in the accessibility of different types of information, and so on. Consequently, because any arguments against the psychological reality of the IH process would have to spell out why these basic elements are peripheral rather than central to the process of generating explanations, we suspect that such arguments face an uphill battle.

Gelman & Meyer suggest that the inherence-based explanations that we hypothesized to stem from the IH actually stem from a more basic assumption that uniformities in the environment are “real, stable, and nonaccidental” (para. 5). On this view, people are predisposed to see patterns as stable and natural, which leads them to search their memories for inherent facts that can rationalize this predisposition. Thus, inherence-based explanations may not be the result of the process we outlined but rather just a by-product of people’s realist assumption about regularities in the world.

To begin, we note an important similarity between the two accounts. Memory retrieval is a part of Gelman & Meyer’s alternative account, just as it is of ours: In most circumstances, people generate inherence-based explanations by retrieving inherent facts from memory. The key difference from our proposal is that, for Gelman & Meyer, the search for relevant information in memory is not triggered directly by the prompt to explain (as it is in the IH process). Rather, there is an intermediary step: the activation of an assumption about the nature of reality, which then prompts a targeted search for inherent facts consistent with this assumption. However, if inherent facts are more likely to be retrieved anyway because of their greater accessibility, then what need is there for an extra assumption or bias that guides our explanatory intuitions toward these facts? That is, it seems unnecessary to invoke a new psychological entity, especially one whose origin and nature are unclear, when the prevalence of inherence-based explanations can be accounted for by a set of non-mysterious cognitive processes, such as those in the IH, whose reality has already been documented. The typical (inherence-based) output of the IH could potentially be described as a sort of assumption, but it would nevertheless be the IH process that is psychologically real.

Are there cases, however, where inherence-based explanations could stem only from a realist assumption? If such cases existed, they could provide reason to endorse this alternative proposal over ours. In that spirit, Gelman & Meyer raise the point that inherence-based explanations are generated even for novel or unfamiliar facts. Because the realist assumption they propose is formulated at an abstract level, it can easily apply to novel patterns. In contrast, since the IH relies on previous knowledge, it might appear that it cannot handle novel patterns because the shotgun would not be able to retrieve much of anything. Or would it? Let us consider the evidence that Gelman & Meyer have in mind here, which suggests that children generate inherence-based explanations for novel patterns such as that snakes have holes in their teeth or that girls are good
at a game called gorp (Cimpian & Markman 2009; 2011). Although these patterns are indeed unfamiliar to the 4-year-olds in these studies, many of the constituents of the patterns are actually familiar (e.g., snakes, teeth, girls, games), so there is quite a bit of information in memory for the shotgun to retrieve and pass on to the storyteller. In other words, even if the expanandum itself is novel, not all its constituents are novel, and thus there is almost always something for the shotgun to retrieve as raw material for the IH.1 The shotgun would probably be thwarted only by complete novelty (e.g., Jabberwocky-style questions such as “why are borogoves minsy?”), but of course a realist assumption would not be of much use here either, since it is impossible to explain such facts. In sum, the IH account has little difficulty accounting for the presence of inherence-based explanations for unfamiliar patterns, and thus the added value of a realist assumption is still in question.

It is also notable that the IH proposal covers a broader range of phenomena than Gelman & Meyer’s claim about a realist assumption. For example, our proposal subsumes the evidence for a correspondence bias—that is, the tendency to attribute instances of behavior to inherent traits more often than is warranted (e.g., Gilbert & Malone 1995; Jones & Harris 1967). Although the accessibility difference between inherent and extrinsic facts may be less dramatic here than in the case of broad patterns, inherent features of the person (or of people in general) are still likely to be easily retrieved by the shotgun, leading to their overuse—all other things being equal (that is, assuming the context [e.g., culture] does not make extrinsic facts more readily available than they would otherwise be). Gelman & Meyer’s realist assumption is explicitly formulated as applying to patterns and regularities, and thus cannot provide an account of inherence-based explanations for instances. In sum, the IH process seems both more parsimonious than Gelman & Meyer’s proposal (because it does not need to appeal to special assumptions of unclear origin) and better able to account for the full range of explanations that appear to be biased toward inherence.

Yzerbyt & Demoulin raise another possibility, suggesting that inherence-based explanations stem from a basic need to see the world as a “coherent and predictable” (para. 4) place that can be brought under one’s control. People seek explanations in terms of inherent features because these explanations fill this need particularly well. Although Yzerbyt & Demoulin intended their proposal as a rival to our own, there are ways in which it could be complementary, rather than contradictory, to it. For example, if Yzerbyt & Demoulin are correct in assuming that inherence-based explanations provide a sense of control over one’s environment (but see Jones 1979), then fulfillment of this basic need could provide a functional reason for the presence of an inherence heuristic in human cognition. We have already argued that the IH process is present in part because it fulfills the drive to understand the world. Yzerbyt & Demoulin’s point suggests that it may also fulfill the (related) drive to control the world.2

At a more mechanistic level, this motivation to see the world as a stable, predictable place may also modulate the operation of the IH process. For example, high levels of this motivation, whether chronic or situation-specific, might enhance the accessibility of inherent facts in memory (e.g., Anderson et al. 1996), leading to a stronger inherence skew in the content of the shotgun; similarly, high levels of this motivation may lead reasoners to endorse the inherence-based output of the IH immediately and without question. Thus, in addition to suggesting a possible reason why people rely on the IH, the motivation to see the world as predictable and controllable may be directly involved in the how of the heuristic—in the specifics of its operation.

However, Yzerbyt & Demoulin’s intention was to propose an alternative process for generating inherence-based explanations, not to simply offer suggestions regarding the process we proposed. Their alternative process follows a course similar to Gelman & Meyer’s: Explanations appealing to inherent features emerge from a memory search that (a) occurs only after the activation of the hypothesized need to see the world as stable and predictable, and (b) targets inherent features specifically, so as to generate explanations that fill this need. The first question we raise here is the one we raised about Gelman & Meyer’s proposal: Is the motivation invoked by Yzerbyt & Demoulin necessary (at a mechanistic level) to produce inherence-based explanations? Most likely, it is not. The process outlined in our proposal, which relies on basic, well-understood cognitive components, is sufficient to generate inherence-based explanations without any motivated influences. Although this heuristic process can be influenced by reasoners’ needs and goals, its output will be biased toward inherence even in the absence of such an intervention. What, then, is to be gained by invoking the need for an orderly, controllable world?

Even though this psychological motive may not be necessary to arrive at inherence-based explanations, perhaps it is sufficient. In that case, we would have to adjudicate between two alternative, individually sufficient paths to inherence-based explanations: the IH process and Yzerbyt & Demoulin’s motivated process. What reasons are there to favor one over the other? Yzerbyt & Demoulin provide two justifications for their preferred mechanism. The first is the occurrence of vague inherence-based explanatory intuitions (e.g., “there is something about girls that explains why they wear pink”). These intuitions are argued to be the initial explanatory products of people’s need for a predictable world, before inherent features are found in memory (or discovered via later learning) that can be used to elaborate these initial intuitions into something less vague. But this point cannot differentiate between the two accounts: As explained in our target article, such vague intuitions are a potential output of the IH process as well. Specifically, they occur when the content retrieved by the shotgun cannot be assembled into a sensible explanation, and as a result reasoners are left with a vague impression that some of the (predominantly inherent) facts activated from memory will ultimately explain the observations. Phenomenologically, this failure of the storyteller to come up with a concrete explanation will often translate into the intuitions Yzerbyt & Demoulin have in mind (e.g., “there is something about girls…”). Thus, the occurrence of these intuitions is not a reason to favor Yzerbyt & Demoulin’s motivated model. The second justification Yzerbyt & Demoulin provided for this model was that, due to confirmation bias, inherence-based intuitions are resistant to change. However, it is unclear to us why the existence of a confirmation bias in human cognition should provide unique support for
Yzerbyt & Demoulin’s motivated model. Once an inherence-based intuition has been generated, even a vague one, confirmation bias will make it resistant to change regardless of the process by which this intuition came about (e.g., Nickerson 1998). Consequently, Yzerbyt & Demoulin’s proposed process cannot derive special support from the general inertia that characterizes human beliefs.

So far, we have argued that Yzerbyt & Demoulin’s reasons for favoring their motivated process over ours may not be convincing. Are there reasons to favor our proposal instead? To begin, we highlight the promise of the IH model via a brief historical analogy. When the correspondence bias was first documented in the late 1960s, one of the accounts offered was that this bias stems from the fundamental need to predict and control the world. Perceiving inherent dispositions in others’ behaviors was thought to satisfy this need by organizing the complex stream of behavioral data into invariant “chunks” (i.e., dispositions) that could then be used to predict and influence future behaviors (see Heider 1958). Even in the early stages of research on the correspondence bias, though, few researchers believed that motivational drives could provide a sufficient account of this bias (e.g., Jones 1979; Ross 1977). Jones (1979), for example, expressed serious reservations about the need-for-control argument: “How can this ubiquitous proneness to attributitional error really facilitate control? Why should we feel more in control when we think we understand a personal disposition than when we think we understand the situational context of behavior?” (p. 116). In part as a result of these doubts, research into the cognitive mechanisms underlying this bias continued unabated until ultimately settling on a model that appeals to fast heuristic processes operating on easily accessible information with loose supervision/correction by working-memory-dependent processes (e.g., Gilbert 2002; Gilbert et al. 1988; 2003; Trope & Gaunt 2000). Motivational factors such as the need for predictability and control are peripheral in contemporary accounts of the correspondence bias. The point of the analogy is this: If heuristic cognitive mechanisms were sufficient to account for the inherent bias in people’s explanations for behavior, it seems plausible that such mechanisms will also be sufficient to account for the more general inheritance bias under investigation here.

Consistent with this possibility, inheritance-based explanations seem to behave exactly how one would expect the output of a heuristic (as opposed to a purely motivational) process to behave. First, these explanations are most prevalent in circumstances where inherent facts are most accessible and thus most likely to be used for generating quick explanations (e.g., Cimpian & Erickson 2012; Cimpian & Markman 2009; 2011). Second, inheritance-based explanations are judged to be more plausible when people’s cognitive resources are taxed by a secondary task (Salomon & Cimpian, in preparation). Again, this is precisely what we would predict: Heuristic responses generally increase under cognitive load (e.g., Epley & Gilovich 2006; Gilbert et al. 1988). Third, inheritance-based explanations are more strongly endorsed by people with lower scores on a fluid-intelligence test (Salomon & Cimpian 2014). The same is true of the output of other intuitive heuristics (e.g., Stanovich & West 2000; see also Stanovich & West 2008). Fourth, inheritance-based explanations are more strongly endorsed by children (Cimpian & Steinberg, in press), a developmental difference that parallels that identified for other heuristics (e.g., Kokis et al. 2002; Toplak et al. 2014). Fifth, inheritance-based explanations are evaluated more positively by people who tend to adopt less reflective or effortful thinking styles (Salomon & Cimpian 2014). The negative relationship between reflective thinking/cognitive styles and reliance on heuristic intuitions is also well established (e.g., Epley & Gilovich 2006; Stanovich & West 2000; Toplak et al. 2014). Together, these data speak to the plausibility of our heuristic model as a source of people’s inheritance-based intuitions; they also suggest that these intuitions are unlikely to be just the by-products of a motivation to see the world as predictable and controllable, as argued by Yzerbyt & Demoulin. (These data may also pose a challenge to Gelman & Meyer’s proposal of a realist assumption. It is unclear how a simple assumption about the nature of reality would be able to account for this sort of systematic variability in the prevalence of inheritance-based explanations.)

To conclude, neither of the alternative sources of inheritance-based intuitions proposed by our commentators (a realist assumption about regularities, and a need for predictability and control) appears to provide a viable alternative to the IH account.

R2. The inheritance heuristic and system justification

In this section, we discuss the potential role of the IH in promoting a tendency to defend the societal status quo. To reiterate, the typical output of this heuristic process consists of explanations that appeal to inherent facts. Regardless of what observations this output is used to explain, it will tend to make these observations seem natural and sensible (rather than contingent or arbitrary). If, for example, someone explains why OJ is for breakfast and appealing to its sleep-chasing tangy taste, then it is only a small further step to assume that this pairing of OJ with breakfast makes perfect sense. It should be apparent that this way of thinking has direct relevance to the phenomenon of system justification. When used to gain an understanding of one’s society, the typical (inheritance-based) output of the IH will tend to make whatever sociopolitical configuration is in place when the IH is triggered seem natural and sensible, and thus deserving of one’s support. Thus, the IH can “fire system-justifying bullets” (Gaucher & Jost, para. 3) for completely non-motivated, non-ideological reasons—simply because inherent facts (or “facts”) are often easily accessible in memory and are thus low-hanging fruit for a process that’s looking for quick-and-dirty explanations. In other words, a motivation to defend the status quo is not necessary for status-quo-defending attitudes to emerge as a result of the IH process. This is not to say that motivational states are never involved in the operation of the IH—they can be, as explained in our target article, and they may thus contribute to the generation of system-justifying outputs. But the important point here is that defense motives are not needed for such outputs to emerge.

Our argument so far is that the IH process is sufficient to give rise to system justification— even in the absence of motivated influences. Given the ubiquity of explanation in
everyday life, we also suspect that the IH is a prolific source of system-justifying cognitions. We do not, however, endorse the more extreme claim that the IH is necessarily involved in any instance of system justification. Psychological states that reinforce the status quo may emerge via other processes as well. For example, exposure to the repeated pairing of certain groups with positively valenced markers of high status (e.g., wealth, power) could result, via simple associative learning, in similarly valenced emotions toward these high-status groups, regardless of one’s own group membership (see Newheiser & Olson). Thus, the IH is a frequent entry point into system justification, but it is not the only one.

We now go on to address some of the points raised in the commentaries. Gaucher & Jost wondered “why – in the absence of social and motivational considerations – the mental shotgun fires system-justifying bullets” (para. 8). The answer is simple: because inherent, and thereby system-justifying, bullets (e.g., stereotypes about the inherent traits of different groups) are at its fingertips (e.g., Devine 1989). In effect, system-justifying explanations are a prime example of how easy heuristic judgments can really miss the mark, with serious consequences. We should clarify, however, that this heuristic framework for understanding system justification does not negate the role of defense motives in this phenomenon, a role that Gaucher & Jost and Uhmann, Zhu, Brescoll, & Newman (Uhmann et al.) illustrate with many convincing examples. Rather, our argument acknowledges the importance of these motives and fits them into a broader account that also spells out some of the possible cognitive underpinnings of people’s tendency to defend the status quo. This account has additional advantages besides the deeper mechanistic understanding it may provide. For example, by highlighting the frequency and ease with which inheritance-based explanations are generated, our proposal provides new insight into why defense motives co-opt these explanations with such remarkable regularity. Another advantage of our heuristic account is that it unifies a number of apparently disparate areas of research. To us, the possibility that far-flung phenomena such as the correspondence bias and system justification are even partially underlain by a simple explanatory heuristic is intellectually exciting and worthy of serious consideration.4

In their commentary, Uhmann et al. endorse the existence of deep links between the IH and system justification but caution against equating one phenomenon with the other. We agree. As clarified earlier, we see the IH as one of multiple pathways to system justification. Other pathways include low-level, implicit associations of the sort described by Newheiser & Olson, as well as motivational influences that don’t operate through the IH process: Although the motivation to defend the status quo can indeed skew the output of the heuristic, it almost certainly leads to system-justifying judgments by means unrelated to the IH as well (e.g., beliefs about karma or fate). The only potential point of disagreement with Uhmann et al. arises with respect to their claim that system justification is necessarily motivated. Because the IH can produce inheritance-based explanations for societal patterns without any motivated prompts, and because the content of these explanations can in and of itself justify existing societal arrangements, Uhmann et al.’s claim that motives are definitional of system justification seems too strong.

Newheiser & Olson add a developmental layer to this discussion by reviewing some of their evidence of system justification in childhood (in particular, South African minority children’s striking outgroup preferences) and by raising interesting questions about how this evidence should be interpreted: Does one need to invoke inheritance-based explanations to account for children’s implicit preference for high-status groups, or would low-level associations suffice? Newheiser & Olson argue that associations may be all that is needed to explain their data, and we agree. However, this is not to say that low-level associations and the ensuing implicit preferences for high-status groups exhaust the full extent of system justification in childhood. As Bigler & Clark remind us, children not only detect regularities in their social world but also endeavor to explain these regularities. If the IH is part of the process by which these regularities are explained (and Bigler & Clark agree that it most likely is), then we would expect to see stronger, more explicit forms of system justification in childhood as well. That is, we would expect children to display not only implicit preferences for high-status groups (of the sort described by Newheiser & Olson) but also explicit cognitions that frame existing societal structures as fair because they are the natural by-products of the relevant groups’ inherent features. Hints in the existing literature suggest that such cognitions may indeed be present: Consider, for instance, children’s tendency to legitimize the dearth of women and people of color in positions of authority by appealing to these groups’ inherent features (e.g., lower intelligence; Bigler et al. 2008), or the fact that 4-year-olds explain even novel facts about social groups (e.g., that girls are good at gorp) in terms of inherent traits (Cimpian & Erickson 2012; Cimpian & Markman 2011). In addition to these clues, direct tests of our account’s developmental predictions are currently under way, with promising results (Hussak & Cimpian 2013). Because some of these tests involve novel social categories, we are also able to rule out the possibility that children’s system-legitimating explanations occur only as a means of rationalizing an already-acquired implicit preference for the high-status groups (as argued by Newheiser & Olson).

To summarize, we argued that the IH is a sufficient – and powerful – source of system-justifying attitudes, with roots that stretch deep into early childhood.

R3. The inheritance heuristic and is–ought inferences

The typical (inheritance-based) output of the IH leads to a tendency to defend the status quo in part because it makes the status quo seem sensible and appropriate – the way things should be. If the tangy taste of OJ is invigorating, then it stands to reason that OJ should be consumed for breakfast (as opposed to, say, dinner). Inferences about shoulds and oughts are even easier to make when the inheritance-based output of the IH already relies on value-laden explanatory notions (e.g., the taste of OJ makes it ideal for breakfast) as opposed to causal notions (see, e.g., sect. 2.2.3 of the target article). For these reasons, we argued that the IH may be one of the psychological mechanisms underlying the frequent transitions from is to ought that Hume (1740/2000) observed in his
contemporaries’ reasoning – transitions that seem no less common today. In this section, we clarify a few aspects of our argument and then discuss a proposal for an alternative source of is–ought inferences (brought up by Bartsch & Estes).

First, it is important to note that our hypothesis concerns human psychology, not metaethical principles: The IH proposal explains why people are prone to take what is the case as a signal of what should be the case. However, this proposal does not commit us to any particular position on the issue of whether moral truths can in fact be legitimately derived from statements of fact. Although we suspect that many is–ought inferences are indefensible (which is why we occasionally used the term is–ought errors in the target article), we do not want to be taken as saying that moral judgments are necessarily divorced from, and altogether a different sort of entity than, empirical judgments (for defenses of naturalistic moral realism, see Boyd 1988, Railton 1986, and many others). We thank Dubljievic & Racine for the prompt to clarify our position on this issue.

We also want to clarify a related matter that came up in Dubljievic & Racine’s commentary: The IH is a psychological process that informs everyday explanatory (and, by extension, moral) reasoning. We are more hesitant to read the operation of the IH into the deliberative reasoning processes that, say, philosophers might rely on in constructing their arguments about what makes something morally virtuous or reprehensible. Thus, we would caution against using examples such as those mentioned by Dubljievic & Racine (e.g., Kant’s deontological arguments) as illustrations of heuristic reasoning. This is not to say that vestiges of intuitive heuristics (including the IH) are never present in scientific or philosophical arguments. We suspect that they are, but differentiating between the products of heuristic and deliberative thought becomes more difficult in these circumstances.

To reiterate, we are arguing that is–ought inferences may stem from the operation of the IH. Bartsch & Estes propose an alternative source for these inferences, particularly as they occur in children’s reasoning. Specifically, Bartsch & Estes argue that children may assume that the world is exactly as it should be because it conforms to human goals and intentions: “If the way things are is the way people want them to be, then maybe it’s all good” (para. 6). We are in complete agreement with Bartsch & Estes regarding the centrality of reasoning about mental states in children’s cognitive life. With respect to explanations per se, however, we argue that such mentalistic reasoning is not as ubiquitous as Bartsch & Estes claim. We first detail how this argument follows from the IH proposal. We then review some of the evidence on this point, which seems to favor our account over Bartsch & Estes’s.

According to the IH proposal, the content of people’s everyday explanations tends to be supplied by shallow memory searches that often start with the main constituents of the explanandum as retrieval cues. For example, if children are trying to explain why OJ is for breakfast, they may start by retrieving any easily accessible facts about OJ and breakfast. Although these most-accessible facts may occasionally involve mental states (e.g., Mom likes OJ), in many circumstances the shotgun will retrieve only facts about the relevant objects’ inherent features (e.g., OJ is sour), which are the sort of facts that dominate our semantic representations (e.g., McRae et al. 1997; 2005). As a result of this retrieval bias, heuristic explanations for facts about the world may actually appeal to human decisions less, rather than more, often than would be normatively warranted – the opposite of Bartsch & Estes’s prediction.

Is there evidence that can adjudicate between these claims? To begin, although Bartsch & Estes invoke childhood artificialism (a hypothesized tendency to see human agency at the origin of most phenomena and events) to support their argument, we note that Piaget’s (1929/1967) conclusions on this topic were contradicted by subsequent work (e.g., Gelman & Kremer 1991). In fact, our reading of the relevant developmental literature points to the conclusion that children typically fail to see the human agency behind many features of their world that in reality are the result of human intentions and decisions. (We are not alone in this interpretation of the literature. For instance, the realist assumption proposed by Gelman & Meyer has a similar flavor.) As an illustration, children often fail to understand that certain aspects of the world, such as modes of dress or word–referent mappings, are due to mutable social conventions (e.g., Brook 1970; Gabennesch 1990; Kalish 1998; Lockhart et al. 1977). Instead, children seem to understand many of these conventions as permanent fixtures of the world that cannot be changed even if social consensus favored such a change (that is, even if people wanted a change). Similar intuitions were also documented in a recent series of studies investigating the developmental course of the IH (Cimpian & Steinberg, in press). As would be expected if the IH were present early in development, preschool-age children often denied that coins are round, fire trucks are red, and so forth, because people wanted them to be so, endorsing instead explanations that appealed to the inherent natures of the relevant objects. In summary, the totality of the developmental evidence suggests to us that an appreciation of human intentions is not as central to children’s explanations of the world as Bartsch & Estes argue. Rather, children tend to assume that the origin of many environmental uniformities can be found within these uniformities themselves (specifically, in the features of their constituents), as predicted by our IH account.

R4. The Inherence Heuristic and Psychological Essentialism

In our target article, we proposed that the IH, with its tendency to output explanations that appeal to inherent facts, could serve as a precursor to psychological essentialism. This was perhaps the riskiest move in our paper, in no small measure because essentialism is a moving target. As was clear from our commentaries, there is no general agreement on what counts as essentialism; and if we don’t know what essentialism is, it is of course impossible for any theory (including ours) to successfully explain how it emerges. In what follows, we provide a brief analysis of the construct of essentialism as we see it reflected in the commentaries. Based on this analysis, we conclude that this construct has become so diluted and broad that it is unlikely to be picking out a single psychological phenomenon. We then clarify which of the phenomena the essentialism label currently refers to we propose to explain with the IH. Finally, we answer the questions that came up
regarding the developmental process by which the IH leads to the emergence of these specific essentialist phenomena.

**R4.1. What is essentialism?**

We share Haslam’s impression that the term essentialism has been applied very liberally in recent years, to the point where it has become hard to interpret and perhaps no longer useful for further theorizing. As an illustration of this issue, next we list the disparate intuitions that our commentators labeled as essentialist. To preview, the only feature these intuitions all seem to share is that they involve an appreciation for a reality beyond the obvious – for the fact that there is more to the world than meets the eye. But do these intuitions necessarily stem from the same psychological process? This is the crucial question if we are to continue studying essentialism as a unitary phenomenon. Unfortunately, it seems to us that the answer is no.” In fact, a variety of different cognitive processes lead people to rely on nonobvious factors, including many whose outputs have not (so far, at least) been included under the essentialism rubric (e.g., reasoning about others’ [nonobvious] mental states). Thus, if the only feature shared by all phenomena labeled as essentialism is an appreciation for the nonobvious, and if this feature can arise from a diversity of sources, it seems hopeless to attempt a unified theory of essentialism. Instead, more progress might be made if we first grouped essentialist intuitions (i.e., intuitions that appeal to nonobvious factors) by their suspected etiology/origin and then investigated each of the resulting groups independently. This sort of classification, however, goes beyond what we can accomplish in the present response. Here, we limit ourselves to making the point about the diverse etiologies of essentialist intuitions and then carving out a portion of these intuitions for the IH to explain.

In this spirit, here are some of the intuitions that were mentioned in the commentaries as instances of essentialism:

**Intuition 1.** A thing’s category membership is given by a nonobvious physical entity present in it. For instance, Fido is a dog, and possesses dog-typical properties, by virtue of some of its DNA. This seems to be the intuition that is most widely classified as essentialist (e.g., Dennehy; Haslam; Hood; Noles & Danovitch; Olivola & Machery). **Intuition 2.** A thing’s unique identity is given by a nonobvious physical entity present in it (Hood). For instance, Fido is the individual it is, with its unique properties, by virtue of some of its DNA. **Intuition 3.** A thing’s unique identity is given by a nonobvious “essential element” that picks out the thing across space and time – perhaps something akin to a mental index that tracks it (Hood, para. 4). For instance, the ship of Theseus maintains its identity over a complete change in its material composition because it retains this essential identity-preserving element. The essence in this intuition is nonphysical (because it is completely independent of the object’s material composition), which contrasts with Intuition 2 (where the essence inheres in the object). **Intuition 4.** A thing’s value (monetary, sentimental, etc.) is given in part by a nonobvious entity obtained via the thing’s relations to other entities (e.g., oneself, a famous person, a historical event; see Hood and Noles & Danovitch). For instance, a pen that we own, a pen that belonged to Beyoncé at some point, and a pen that was used to sign the U.S. Constitution are all endowed with some nonobvious entity (an essence) that gives them a “special status” (Noles & Danovitch, para. 3) and makes them more valuable than pens that are molecule-for-molecule identical. This is the essentialist intuition often invoked in discussions of ownership and authenticity (e.g., Gelman 2013; Hood & Bloom 2008). Essences account for the value of an object here rather than for its category membership (Intuition 1), physical features (Intuitions 1 and 2), or unique identity (Intuitions 2 and 3). These value-conferring essences are acquired or lost over the course of the object’s history, depending on the presence of the relevant relations (e.g., whether the object is in one’s possession). This fluidity contrasts with Intuitions 1–3, where the essences seemed more than just temporary attributes that an object can gain or lose with the circumstances. Also of note, the essences involved here need not be physical. When a person assumes ownership of an object, for example, the object appears to acquire whatever essential entity that gives it its special status even without any physical contact with its new owner (and thus presumably without the possibility of any physical changes occurring to it). **Intuition 5.** There is a deeper, nonobvious reality behind the surface data (Bookstein). This seems to be the cognitive impulse behind much scientific activity, perhaps even down to the level of the statistical methods, as Bookstein points out. For example, the logic behind something as simple as calculating the mean of a number of observations seems to presuppose the existence of a true value of which these observations are indicative. This essentialist intuition places almost no constraint on what the deeper reality (or the essence) might be – just that it exists and that it is the source of the observed data.

These are just the examples of essentialism that came up in our commentaries – more can be found in the broader literature, but the present set is sufficient to make our point: Although using the term essentialism to refer to all of these intuitions gives the impression that they are all outputs of the same psychological process, it is hard to see how that could be the case. Given what is known about our species’ neurocognitive architecture, it seems implausible, for example, that a process that tracks object identity across space and time (Intuition 3) would be one and the same as a process that computes the subjective reward value of “emotional objects” (Hood, para. 4; Intuition 4) or a process that provides causal explanations for a category’s features (Intuition 1); these judgments all involve different types of computations that are likely to be instantiated in different brain substrates. Thus, rather than forming a coherent essentialism category, judgments that assume the presence of nonobvious entities appear to arise via a range of different mechanisms. Some of these judgments (such as Intuition 5 about the presence of a deeper reality) might stem from the basic human motivation to understand and explain what is observed (e.g., Callanan & Oakes 1992; Murphy & Medin 1985; Schulz...
R4.2. How might the inherence heuristic lead to the development of (Intuition 1) essentialism?

We proposed that Intuition 1 (which applies to natural and social kinds) may emerge as an elaboration of the IH’s output over the first few years of life. Briefly, this hypothesized transition might occur as follows: Initially, children invoke the IH to explain the uniformities observed across the members of a category. In many cases, however, children may not be able to assemble the inherent facts pulled up by the shotgun into coherent explanations; as a result, children may be left with a vague impression that some of these inherent facts will ultimately be sufficient to explain the uniformities in question. At this point, children haven’t yet developed Intuition 1 because they are not conceiving of the characteristics of the relevant categories as emerging from a single, internal source. However, the unresolved, to-be-determined output of the IH may in fact account for other essentialist intuitions, particularly ones that impose fewer constraints on the location and nature of the explanatory “essences.” For example, this early output of the IH might be exactly what Strewness glosses as a belief in the “causal efficacy of category membership” (para 7). In other words, the heuristic impression that some inherent facts about the members of the category will eventually explain their characteristics could easily be mistaken for, and thus described as, a belief in the causal efficacy of category membership per se. This is only a conjecture at this point, though, so it will be important to test empirically whether we are correct; whether what Strewness describes as an explanatory appeal to the brute fact of category membership (e.g., the mere fact of being dogs causes dogs to bark) is actually an appeal to inherent facts about the category (e.g., some features of dogs cause them to bark).

So how does Intuition 1 emerge out of the early intuitions supplied by the IH? The unresolved nature of many of the heuristics outputs over the first few years of life may prompt children to look for plausible means of arriving at more specific (and perhaps more satisfying) explanations. We argued that children could use at least two sources of early causal knowledge to elaborate the vague output of the IH, both of which would also bring this output closer in content to Intuition 1. In particular, children could rely on their preexisting causal beliefs about insides (e.g., helping vs. hindering) or internal energy (e.g., Gottfried & Gelman 2005; Hatano & Inagaki 1994; Morris et al. 2000). Both of these are broad causal principles that help children reason about the basic functioning of living things (e.g., their movement, their growth). However, when used as a means of refining the output of the IH with respect to why the members of a particular category have the features they do, these causal forces will be translated into category-specific versions of their general forms. For example, even though children might at first conceive of insides as the completely category-general causal force behind animal movement (with no distinctions made between the insides of dogs, cats, birds, etc.), when children later use this belief in their attempt to determine which inherent fact(s) about dogs might explain their characteristic features, it will necessarily take on a category-specific form (that is, children will need to appeal to dog insides and their unique causal powers). Our claim is that, by way of such refinement processes, children gradually arrive at intuitions that can be described as Intuition 1.

Several commentators had questions about this elaboration process. For example, Marmodoro et al. wondered, among other things, what motivates the transition to essences (is it something that children find “rationally compelling” [para. 5], or is it learned?), what is gained by an appeal to essences, and why children should assume that a single essential property explains the rest of a category’s properties. On our account, the transition to (Intuition 1) essentialism is at least partially driven by a motivation to resolve the ambiguity caused by the storyteller’s inchoate output and thus to find a more satisfying explanation for the uniformities observed within a category. Elaborating this ambiguous output with the help of preexisting notions about insides and/or energy may in fact be a rationally compelling (or at least defensible) means of identifying a more specific, and thereby more satisfying, explanation. Thus, we suspect that essentialist beliefs are in large part constructed rather than learned, especially since young children encounter very little talk about essences or essencelike causes (Gelman et al. 1998; but see Gelman et al. 2010; Rhodes et al. 2012). (Of course, later learning could alter the content of these early essentialist beliefs, such as when children start learning about genetics.) As the preceding discussion suggests, what children gain by transitioning over to Intuition 1 is (what appears to them to be) a more satisfying explanation than the one that the IH was able to provide. Finally, note that the emphasis on a single essential feature is simply a by-product of this elaboration process, not a goal of it. Because notions about insides and/or energy are available to guide children’s search for a more specific explanation, and because these
notions rely on a single causally efficacious entity, children end up with beliefs about a single essential feature causing many others.

Kinzler & Sullivan asked whether the elaboration process we are proposing involves true conceptual change or simply the enrichment of existing conceptual content. Given the cognitive leaps involved in (a) bringing prior beliefs about insides and/or energy to bear on the output of the IH, and in (b) creatively twisting these category-general causal forces into category-specific explanatory roles, we might classify this process as an instance of conceptual change. Point (b) is also relevant to a comment by Bastian, who questioned whether internal energy, which is a general and fluid causal force that does not distinguish between kinds, could be involved in the process by which kind-specific essences emerge. We agree with Bastian about the category-general nature of internal energy and other presumed vitalistic processes. In fact, this is precisely the reason why we emphasized that children’s beliefs about internal energy would have to be twisted or distorted to fit the mold of an esseccelike inherent feature. This may be a big conceptual leap for children, but the considerable causal power usually assigned to this energy makes it a very appealing means of refining the output of the IH into a plausible story.

Finally, several commentators pointed to potential social influences on the content and endorsement of essentialist beliefs (O’Connor & Joffe; Olivola & Machery; Rhodes). As detailed throughout this response, the operation of the IH is sensitive to such sociocultural factors. As a result, our account is in principle capable of accounting for their influence on the heuristic’s offshoots.

As a final note, we shamelessly highlight Haslam’s assessment of the present proposal as a “useful and potentially generative idea” that “deserves serious consideration by psychologists who study essentialist thinking” (para. 8). Naturally, we share Haslam’s optimism. And, since the bulk of empirical work is ahead of us, we hope that the answers provided here have clarified our theoretical position to a point where others may be enticed to join us in testing its predictions.

NOTES
1. There is no guarantee, however, that the storyteller will be able to shape the content at its disposal into an explanation. Whenever it fails to do so, the output of the IH will be the vaguer, “to-be-determined” sort of intuition – namely, that some of the facts retrieved from memory are relevant to an explanation in a to-be-determined way.
2. At this point, the most we can say is that the IH fulfills these needs by providing a subjective sense of understanding and control. Whether the IH truly enhances people’s understanding and control of the environment depends on further questions about its normative accuracy.
3. Yzerbyt & Demoulin asked whether the IH scale that we described in our target article might measure people’s adherence to social norms rather than a reliance on inference-based explanations. Although this scale correlates with scales measuring people’s adherence to social norms (as our account would predict as well), we have evidence that it also taps directly into people’s explanations of the world. For example, in a construct validation study described by Salomon and Cimpian (2014), we asked people to justify why they agree or disagree with our scale items. As we would expect, agreement with our scale items was overwhelmingly justified with inherent reasons (e.g., traffic lights use red to signal stop because “red is a warning signal in nature”). Conversely, disagreement with the scale items was justified with extrinsic reasons (e.g., “we have learned that red means stop so I feel this is why red means stop”).
4. As a note of historical interest, we are not the first to posit deeper commonalities between the correspondence bias and system justification. In the last paragraph of their classic article on the fundamental attribution error (which we refer to as correspondence bias here), Ross et al. (1977) speculate that “this distortion in social judgment [i.e., the bias to attribute behavior to corresponding inherent traits] could provide a particularly insidious brake upon social mobility, whereby the disadvantaged and powerless overestimate the capabilities of the powerful who, in turn, inappropriately deem members of their own caste well-suited to their particular leadership tasks” (p. 494).
5. The IH might also be at the source of intuitions similar to Intuition 2 (about the physical essences of unique individuals), especially since the IH is also triggered to explain specific features or behaviors. We leave that argument for another occasion, however.
6. We should also reiterate that inherent is not synonymous with internal, and thus that Streven’s arguments against what he calls insides essentialism miss their target when directed at our proposal.

References

References/Cimpian & Salomon: The inherent heuristic

Gelman, S. A. & Bloom, P. (2000) Young children are sensitive to how an object was created when deciding what to name it. Cognition 76:91–103. [aAC]
References/Cimpian & Salomon: The inheritance heuristic


