Thinking Like A Scientist: Innateness as a Case Study

(Forthcoming in *Cognition*)

Joshua Knobe[1,2] and Richard Samuels[3]

[1]*Program in Cognitive Science, Yale University,* [2]*Department of Philosophy, Yale University,* [3]*Department of Philosophy, Ohio State University*

Abstract: The concept of innateness appears in systematic research within cognitive science, but it also appears in less systematic modes of thought that long predate the scientific study of the mind. The present studies therefore explore the relationship between the properly scientific uses of this concept and its role in ordinary folk understanding. Studies 1-4 examined the judgments of people with no specific training in cognitive science. Results showed (a) that judgments about whether a trait was innate were not affected by whether or not the trait was learned, but (b) such judgments were impacted by *moral* considerations. Study 5 looked at the judgments of both non-scientists and scientists, in conditions that encouraged either thinking about individual cases or thinking about certain general principles. In the case-based condition, both non-scientists and scientists showed an impact of moral considerations but little impact of learning. In the principled condition, both non-scientists and scientists showed an impact of learning but little impact of moral considerations. These results suggest that both non-scientists and scientists are drawn to a conception of innateness that differs from the one at work in contemporary scientific research but that they are also both capable of 'filtering out' their initial intuitions and using a more scientific approach.

Many of the concepts that appear quite frequently in scientific research have no counterparts in ordinary folk understanding. There is no folk concept of the Higgs boson, or of the lateral geniculate nucleus, or of Calabi-Yau manifolds. These concepts were invented by scientists, and they cannot be understood outside the context of the scientific theories in which they originally appeared.

The concept of innateness, however, seems importantly different. This concept plays a prominent role in research within cognitive science, but it was not originally developed by scientists. On the contrary, discussions of innateness appear in works of both Western philosophy (Cowie, 1999) and Chinese philosophy (Fung Yu-lan, 1953; Wong, 2012) that long predate an empirical science of the mind. Moreover, the notion of innateness appears to have a place in people's ordinary folk understanding. Even if many people have rarely come across the actual English word 'innate,' they can easily understand the claim that certain capacities are 'just built into us,' 'in our genes' or 'in our nature' – notions that are widely held to be related in some way to the scientific concept of innateness (Griffiths, 2002).

For this reason, the concept of innateness has the potential to serve as an interesting case study in the cognitive-scientific study of the distinction between folk and scientific thinking. It affords us an opportunity to look at the ways in which scientists can appropriate a concept from ordinary folk thought but then use that concept for distinctively scientific purposes.

1. Differences between scientific and folk understanding

Existing research has examined a number of ways in which scientific thinking differs from ordinary folk thought. First of all, there are numerous differences in individual domains. Contemporary scientific thinking adopts an inertial theory of motion rather than an impetus theory (McCloskey, 1983), a kinetic theory of heat rather than a substance-based theory (Slotta & Chi, 2006), a variation-based theory of evolution rather than a theory based on species essence (Shtulman, 2006). But not all of the differences are confined in this way to individual domains. There also appear to be broader differences that can be found across a number of different areas of scientific research. Scientists appear to be less content with shallow explanations (Weisberg, Keil,

Goodstein, Rawson & Gray, 2008), to reason less teleologically (Kelemen, Rottman & Seston, 2012), to more carefully separate theory and evidence (Kuhn, 1989; Kuhn, Schauble & Garcia-Mila, 1992), and to use formal tools of statistical inference that sometimes allow them to escape the influence of the biases that so often plague people's ordinary judgments (Kahneman, 2011).

Our focus here will be on another of these domain-general differences. Much of folk thought appears to be influenced by *value judgments* in a way that scientific thinking typically is not. Indeed, recent studies seem to indicate that value judgments exert a surprisingly pervasive impact on ordinary folk cognition.

To take one example, consider the way that people ordinarily decide whether an agent has performed a behavior 'intentionally.' It might initially appear that people's answers to this question should be determined entirely by their beliefs about the agent's mental states (what she believes, what she wants, etc.), but a series of studies appear to indicate that something more is actually involved. People's intuitions about whether a behavior was performed intentionally can actually be influenced by their value judgments (Ditto, Pizarro & Tannenbaum, 2009; Knobe, 2003; Nichols & Ulatowski, 2007; Young, Cushman, Adolphs, Tranel & Hauser, 2006; but see Machery, 2006; Sripada & Konrath, 2011). Thus, suppose that a corporate executive decides to implement a policy and thinks: 'I know that this policy will bring about outcome *x*, but I don't care at all about that. I just want to implement the policy for some other reason.' Did the agent then bring about outcome *x* intentionally? In cases like this, people's intuitions appear to depend on their value judgments. If they believe that outcome *x* is morally bad, they tend to say that the agent brought it about intentionally, whereas if they believe that outcome *x* is morally good, they tend to say that she brought it about unintentionally (Ditto et al., 2009; Knobe, 2003; Leslie et al., 2006; Nichols & Ulatowski, 2007; Young et al., 2006; Zalla & Leboyer, 2011).

Similarly, suppose people are wondering which of the factors that were necessary for a given outcome count as 'causes' and which were merely 'background conditions.' Here again, studies show an impact of value judgments, with morally bad behaviors being classified more as causes and morally good behaviors more as background conditions (Alicke, 2000; Hitchcock & Knobe, forthcoming). Related effects have been

observed for people's use of numerous other concepts, including the concepts of knowledge (Beebe & Buckwalter, forthcoming), freedom (Phillips & Knobe, 2009; Young & Phillips, 2011) and the distinction between doing and allowing (Cushman, Knobe & Sinnott-Armstrong, 2008). These various effects appear to be deeply similar, and it seems plausible that they all have the same underlying cause (Knobe, 2010).

This pattern of intuitions appears to be deeply antithetical to what one would expect to find in a systematic scientific inquiry. Of course, there are difficult questions about the role of value judgments in scientific inquiry, and different scientists may adopt different views about these questions (Douglas, 2009). Still, it seems that few scientists would accept the kind of pattern we find in folk intuitions. Suppose that a scientist announced: 'I have a new theory about the nature of intention. According to this theory, the only way to know whether someone intended to bring about a particular effect is to decide whether this effect truly is morally good or morally bad.' Such a proposal, we predict, would be widely rejected as a framework for scientific research. If two scientists agree about all the purely descriptive facts in a given case, and disagree only about the moral significance of those facts, it seems that these two scientists should not thereby end up disagreeing about any purely scientific questions.

In existing work on these issues, it has been widely assumed that the phenomenon in need of explanation is the surprising role of value judgments in ordinary folk thinking, while the absence of such value judgments from scientific inquiry has been more or less taken as a given (e.g., Knobe, 2010; Levy, 2010). Thus most of the research on these phenomena has been devoted to trying to understand the cognitive processes that allow value judgments to influence ordinary folk thinking. Work in this area has led to the development of a number of competing theories (Adams & Steadman, 2004; Alicke, Rose & Bloom, 2011; Knobe, 2010; Sripada & Konrath, 2011; Uttich & Lombrozo, 2010), but despite extensive empirical study, no clear consensus has emerged.

The present paper will not contribute to this existing debate. Instead, we will be focusing on the converse question. What are the cognitive processes that make it possible for scientific thinking *not* to be influenced by these value judgments?

2. The concept of innateness

The concept of innateness is an especially promising domain for studying this question. Unlike many other concepts that are used in the sciences, the concept of innateness was not invented by scientists. It is firmly rooted in people's folk understanding, and a question therefore arises as to how the use of the concepts within the sciences might differ from its use in ordinary folk cognition.

Existing research on the notion of innateness has examined its application in both scientific and non-scientific contexts. Research on its application in scientific contexts typically proceeds by looking at the uses of 'innate' and its cognates in the scientific literature, and seeking to develop an analysis that makes sense of these uses (e.g., Cowie, 1999; Mameli and Bateson, 2006; Samuels, 2002). By contrast, research on the folk notion of innateness proceeds through systematic experimental studies of people's intuitions (Griffiths, Machery & Linquist, 2009; Linquist, Machery, Giffiths & Stotz, 2011).

The scientific concept(s) of innateness. Although the concept of innateness continues to play a central role in research within cognitive science, it has proven remarkably difficult to spell out explicitly what it means for a trait to be innate. Much of the difficulty stems from what might be called the Interactionist Consensus. It is widely agreed that all traits are the result of a complex interaction between genetic and environmental factors (Kitcher, 1996; Lehrman, 1970). But if all traits are in part the product of the environment, then what could it mean to say that some of these traits count as innate? Researchers have responded to this question by developing a wide array of analyses of innateness (Ariew, 1996; Khalidi, 2002, 2007; Mallon & Weinberg, 2006; Samuels, 2004; for a review, see Griffiths, 2009). These rival analyses differ from each other in important respects, and each of them involves complex theoretical issues that we cannot do justice to here. Nevertheless, it will be helpful to sketch two features, widely regarded as characteristic of innate traits, since they will prove relevant to later sections of the present paper.

First, it is often suggested that a trait that is innate will arise under *all normal environmental conditions* (Sober, 1999). Thus, suppose that a given trait will only arise in

human beings if they receive adequate oxygen. The development of such a trait would require a certain contribution from the environment, but since it is normal for humans to receive adequate oxygen, this environmental contribution would not entail that the trait was not innate. By contrast, suppose that a trait only arises in humans when they undergo brain injuries that lead to lesions in the ventromedial prefrontal cortex. Since such injuries are not a normal part of human development, the environmental contribution here actually would entail that the trait was not innate. A difficult question now arises about what it means for a particular environmental condition to be 'normal.' Different analyses address this question in different ways (Kitcher, 1996), and the issue remains very much an open one.

In any case, it should be clear that this first condition is not sufficient for a trait to count as innate. If we assume, for example, that it is normal for human beings to see the sun, we might conclude that under all normal environmental conditions human beings will acquire a belief that the sun is bright. Yet this belief would not thereby count as innate. There is also something deeply important about *how* a trait is acquired. Once again, this question takes us into controversial territory, and different analyses of the concept of innateness will spell out the relevant condition in different ways (Samuels, 2004). For present purposes, however, we do not propose to take sides. Rather we merely rely on a simple truism that should be entirely uncontroversial. The truism is that innateness is in some way opposed to *learning*. All psychological traits presumably arise in part as a result of environmental influence. But if an organism acquires a certain capacity or item of knowledge by perceiving its environment and engaging in some paradigmatic form of learning – such as classical conditioning or inductive inference – then the resulting trait will not count as innate (Mameli & Bateson, 2011; Samuels, 2002). We take it that this assumption about the scientific notion of innateness should be uncontentious. Indeed, to our knowledge, no hypothesis in psychology – or any related science, for that matter – has ever posited innate traits that are learned.

As one illustration of the role that the opposition between learning and innateness plays in the sciences, consider the debate within the theory-of-mind literature over how children come to have an understanding of false beliefs. One central component of this debate concerns whether or not the concept BELIEF is innate. Some researchers argue

that this concept is learned (e.g., Gopnik & Wellman, 1992), while others argue that it is innate (e.g., Scholl & Leslie, 1999). Presumably, neither side in this debate is committed to the absurd claim that the environment plays no role at all in the acquisition of the belief concept. The question is simply about the precise nature of this role. Researchers on one side claim that children actually acquire the concept by learning from their environment, while those on the other deny that this is the case.

The folk concept of innateness. The concept of innateness figures prominently in contemporary cognitive science, but the concept can also be used by people who have no relevant scientific training. This folk understanding of innateness is an interesting phenomenon, which can be studied in its own right.

Griffiths and colleagues (2009) argue that the folk notion of innateness forms a part of people's folk biology. In particular, the suggestion is that this notion taps into people's folk-biological essentialism (Gelman, 2003; Keil, 1992) – that a trait is classified as 'innate' if it seen as part of an organism's species *essence*. The notion of an essence has a long and checkered intellectual history in which it has been defined many times over. But a trait that is part of a species essence, in the sense intended by Griffiths et al., is one that exhibits three main features:

- Fixity: The trait is hard to change in that its development is insensitive to environmental variation.
- Typicality: The trait is part of what it is to be an organism of that kind. Roughly: Every member of the species, except highly atypical members, has it.
- Teleology: The trait does not merely happen to arise in typical members but actually fulfills a purpose or end for the organism.

According to Griffiths et al., then, we should expect people to judge a trait innate to the extent that they judge it to be fixed, typical and teleological. A series of experimental studies show that people's intuitions about innateness do indeed follow the patterns one would predict on this hypothesis (Griffiths et al., 2009).

Two key features of the Griffiths et al.'s hypothesis require further comment. The first feature concerns the extent to which the folk notion of innateness will be influenced by scientific knowledge. According to Griffiths et al.'s hypothesis, the folk notion of innateness is more or less independent of empirical developments within cognitive

science. For example, the basic criteria associated with this notion are not informed by debates in linguistics or psychology. Indeed, it may have little or no connection to any of the issues discussed in contemporary scientific work. Instead, the folk notion is shaped by more general facts about human cognition, and the best way to understand it is by further examining the nature of naïve essentialism and the character of folk biology.

The second key feature of Griffiths et al.'s hypothesis concerns one specific respect in which we should expect folk and scientific judgments of innateness to diverge. According to their hypothesis, traits are judged innate to the extent that they exhibit the three central features of a species essence (fixity, typicality, teleology). Moreover, this should be so even if one of the traits is learned and the other not. In contrast, if our earlier comments regarding the scientific usage of 'innate' are correct, then we should expect that facts about whether such traits are learned would exert a large effect on scientific judgments of innateness.

3. Hypotheses

With this background in place, we can now introduce a series of predictions and hypotheses.

First, we predict that folk intuitions about innateness will be impacted by their value judgments. Whatever it is that allows value judgments to impact intuitions about other matters (intention, causation, freedom, etc.), we predict that this same process will allow value judgments to impact intuitions about innateness as well.

Second, we predict that scientific judgments about innateness (a) will be impacted by judgments about learning and (b) will not be impacted by judgments about value. Thus, suppose that scientists have an opportunity to see two different cases, presented side by side, that differ only in one obvious respect. If the difference is that one case involves learning and the other does not, scientists will often make correspondingly different judgments about whether the trait is innate. However, if the difference is that one case involves something immoral and the other does not, they will not conclude that this difference is relevant to whether the trait is innate.

These first two hypotheses predict a systematic difference between ordinary folk uses of the concept of innateness and the use of this concept in more formal scientific

contexts. They therefore raise an important question about the cognitive processes that make it possible for the scientific use to depart from the ordinary use. Our discussion of this question will focus on three hypotheses, all of which appear to be antecedently plausible.

On the first, the *overwriting hypothesis*, scientific training leads to the elimination of the folk concept innateness and its replacement by a scientific concept. This new concept of innateness differs in certain ways from the folk concept. In particular, this new concept emphasizes the distinction between learned and non-learned capacities, and it has no place for any considerations of value. Hence, when scientists answer questions about innateness, they never need to explicitly think: 'Be sure not to allow your judgments to be influenced by your values.' Instead, the absence of value judgment arises, as it were, automatically. Scientists make use of a distinctively scientific concept of innateness, and since value judgments play no role in this concept, there is no impact of value judgments on their innateness judgments.

The overwriting hypothesis should be contrasted with what we call 'overriding' hypotheses (c.f. Goldberg & Thompson-Schill, 2009; Lombrozo, Kelemen & Zaitchick, 2007; Shtulman & Valcarcel, 2012). In contrast to the overwriting hypothesis, overriding hypotheses do not propose that scientific training leads to the elimination of a folk concept innateness. Instead, the idea is that more scientific patterns of thinking about innateness results in some way from the supplementation of characteristic folk thinking. There are two versions of this proposal that we distinguish here: the conceptual addition hypothesis and the filtering hypothesis.

On the *conceptual addition* hypothesis, more scientific patterns of innateness judgments depend in part on the acquisition of a new scientific concept of innateness. This proposal is similar to the overwriting hypothesis in that both suggest that scientists acquire a new innateness concept. But whereas the overwriting hypothesis involves the elimination of the folk concept, the present proposal suggests that scientists continue to hold on to the folk concept but also acquire a scientific one that they are able to use under appropriate conditions – such as those that obtain when doing science.

The main reason for thinking that this sort of process might be occurring in the case of innateness judgments is that existing research provides such strong support for the

claim that it is occurring in other domains. A number of studies have examined the judgments of scientists about questions in physics and biology. For each of the questions examined in these studies, it is clear that scientists have developed a theory that goes against people's ordinary folk intuitions. Yet, when participants are asked to answer these questions under highly speeded conditions, even scientists show a tendency to give answers that conform to the folk view (Goldberg & Thompson-Schill, 2009; Keleman, Rottman & Seston, 2012; Shtulman & Valcarcel, 2012). These data suggest that a process of conceptual addition is at work in certain other areas of physics and biology, and one might well think that the same is occurring for the concept of innateness.

Finally, on the *filtering* hypothesis, scientists never actually acquire a distinctively scientific concept of innateness. Instead, they continue to use the folk concept. However, scientists do not merely have a concept that enables them to arrive at judgments about individual cases; they also have certain general principles about which considerations are relevant to innateness judgments. If they see that a pattern of judgments would violate these general principles, this pattern of judgments will be 'filtered out' and a different pattern will be used in its place. In other words, on the filtering hypothesis, scientists never acquire a concept of innateness in which value judgments play no role. Rather, they continue to have a concept in which value judgments do play some role,[1] but they also adhere to a general principle that says 'Do not allow your judgments about innateness to be affected by your value judgments.' When they see explicitly that their judgments are violating this principle, they reject these judgments and try to answer the question in a way that shows no influence of values.

Note that the conceptual addition hypothesis and the filtering hypothesis make quite different claims about the way in which scientists are able to avoid the impact of value judgments. On the conceptual addition hypothesis, scientists have a distinct and purely scientific method for making judgments of innateness. As long as they have the time to reflect carefully on a given case and use this scientific approach (while avoiding the influence of their folk concept), they will be able to answer the question using a method in which moral considerations do not play any role. By contrast, on the filtering hypothesis, scientists never acquire a distinct and purely scientific method for making innateness judgments. Hence, there is no way in which they can avoid the impact of value

judgments simply by applying a scientific approach in which moral considerations never play any role in the first place. The only way for them to avoid the impact of value judgments is to actually think explicitly about which considerations are relevant and apply the principle that innateness judgments should not be affected by value judgments.

Perhaps a simple way to see the difference between the above three hypotheses is by introducing an analogous case involving physical instruments. Suppose we have a cash register that malfunctions in such a way that it adds $1,000 to the bill whenever someone tries to buy an avocado. To address this problem, we might choose any of the following possible approaches. One would be to throw out our existing cash register and purchase a new one that does not malfunction in this way (overwriting). A second approach would be to keep using our original cash register, but also buy a new one, which we could use especially under those circumstances when getting the right answer really matters (conceptual addition). A final approach would be not to get a new cash register at all, but just to keep using the old one and, when we see explicitly that someone is trying to purchase an avocado, to adjust the bill accordingly (filtering). Of course, these approaches might yield exactly the same answers in many cases. However, the approaches themselves are quite different, and one might well ask which of the three is actually at work in the judgments of a particular individual.

To investigate these issues, we proceed in stages. We first conduct quick preliminary studies to see whether people's intuitions about innateness can be influenced by their value judgments (Experiments 1-3) and whether these intuitions are sensitive to the difference between understanding that is learned vs. non-learned (Experiment 4). Then, in Experiment 5, we conduct a large-scale study that examines the judgments of people with no relevant scientific background to the judgments of trained researchers in the field.

Experiment 1

Before we look at the ways in which scientific judgments depart from folk judgments, we need to get a better grasp on the contours of the folk judgments themselves. We begin by examining the impact of moral considerations. The basic strategy here is to use almost exactly the same paradigm that has been used in studies of

intentional action (e.g., Knobe, 2003) but to apply this paradigm to the study of judgments about innateness.

Method

Participants. Forty-nine students volunteered to fill out a questionnaire in the Yale University dining hall in exchange for $1.

Stimuli and procedure. Participants were assigned either to the abilities condition or to the disabilities condition. Each participant then read one or the other version of the following vignette:

> A baby was born with a rare genetic condition. The doctors told the baby's parents: 'If this baby drinks its mother's milk during its first two weeks of life, it will grow up to have extraordinary mental abilities that make it able to solve very complicated math problems [serious psychological disabilities that make it unable to solve even very simple math problems]. However, if you instead give it this expensive formula we sometimes use, it won't develop the extraordinary abilities and will just be normal.'
>
> The parents said: 'We have decided not to give the baby the expensive formula. We will just be feeding it with its mother's milk.'
>
> As expected, the baby grew up to have extraordinary mental abilities that made it able to solve very complicated math problems [serious psychological disabilities that make it unable to solve even very simple math problems].

After reading this vignette, participants were asked whether they agreed or disagreed with the sentence: 'The baby's extraordinary mental abilities [psychological disabilities] were innate.' Participants marked their answer on a scale from 1 ('disagree') to 7 ('agree').

Results and Discussion

Participants gave higher innateness ratings when the parents' action led to special abilities ($M = 4.7$, $SD = 1.9$) than when it led to disabilities ($M = 3.3$, $SD = 1.7$), $t\,(47) =$

2.8, $p < .01$. This result provides at least some initial evidence that people's moral judgments can influence their intuitions about innateness.

Still, one might worry that the results of this first study are susceptible to an alternative interpretation. Perhaps the difference between conditions does not reflect any general impact of moral judgment on people's cognition. Instead, the effect might have arisen as the result of certain relatively straightforward beliefs about which traits are most likely to be innate. (For example, people might simply believe that abilities are more likely to be innate than disabilities are.) To rule out this alternative interpretation, we conducted a second study.

Experiment 2

In this second study, participants were given no information about the nature of the trait itself. They were told only about the genetic and environmental factors that caused the trait to arise. The prediction then was that people would be more inclined to regard the trait as innate when the environmental factors were morally good than when they were morally bad.

Method

Participants. Twenty students volunteered to fill out a questionnaire in the Yale University dining hall in exchange for $1.

Participants were assigned either to the decent treatment condition or to the bad treatment condition. Each participant then read one or the other version of the following vignette:

> Imagine that scientists are trying to understand how people develop a particular trait, which they have come to call Trait X.
>
> The scientists have discovered a surprising fact about people's genes. They have discovered that people's genes work in such a way that almost everyone will end up developing Trait X. In fact, it turns out that children develop Trait X as long as their parents sometimes offer them at least a decent level of treatment [treat them badly].

> Now, just about everyone's parents offer them at least a decent level of treatment [treat them badly] at least sometimes. So, given the way people's genes work, just about everyone actually does develop Trait X.

After reading this vignette, participants were asked whether they agreed or disagreed with the sentence: 'Trait X is innate.' Participants marked their answer on a scale from 1 ('disagree') to 7 ('agree').

Results and Discussion

Participants were more inclined to rate the trait as innate when it was the product of being treated decently ($M = 4.6$, $SD = 1.9$) than when it was the product of being treated badly ($M = 2.7$, $SD = 1.9$), $t\,(18) = 2.2$, $p < .05$. This result lends further support to the view that people's moral judgments are impacting their intuitions about innateness.

The effect observed here seems closely related to the effects of moral judgment that existing studies have found for other concepts (intention, causation, freedom, etc.). Hence, what we are seeing here is presumably just one symptom of a far more general process, and whichever theory turns out to be correct about the general process will probably be correct about the effect observed here as well.

For concreteness, it might be helpful to look briefly at one specific example of a broader theory. Consider the theory according to which people's moral judgments affect their intuitions about various topics because people's moral judgments have an influence on *which possibilities they regard as relevant* (Knobe, 2010). This broader theory provides a natural explanation of the results obtained here. In existing work on both the scientific concept of innateness and the folk concept of innateness, one finds the idea that a trait will only be regarded as innate if it would arise under all relevant environmental conditions (Griffiths et al., 2009; Samuels, 2004; Sober, 1999), but it has proven quite difficult to say precisely how to determine which possible environmental conditions are 'relevant' and which are not (Kitcher, 1996). One obvious hypothesis, then, is that people's moral judgments are impacting their intuitions about innateness by impacting their judgments about which possible environmental conditions are the relevant ones. The possibility that a person's parents would not, even occasionally, treat her decently seems so far-fetched that participants simply ignore it. As far as they are concerned, even if a

trait would not arise under these bizarre conditions, it is still fair to say that the trait is innate. But things begin to look very different when we switch over to the possibility of a person's parents never treating her badly. Although this possibility is highly unusual from a statistical perspective, it does not seem reasonable to ignore it completely. The fact that people regard this possibility as a moral ideal might lead them to continue to see it as highly relevant. For that reason, they might think that a trait that would not arise under these conditions could not be innate.

This theoretical approach has been a controversial one in the case of causation (e.g., Menzies, 2010; Sytsma, Livengood & Rose, forthcoming) and in the various other areas in which it has been applied (Sripada & Konrath, 2011; Uttich & Lombrozo, 2011), and it is bound to be controversial in the case of innateness as well. Our aim here, however, is not to resolve these controversies. Instead, we will be focusing on the question as to how the impact of moral judgments on judgments of innateness might be prevented in a scientific context.

Experiment 3

Experiments 1 and 2 show that people's moral judgments can sometimes impact their judgments of innateness. A question now arises about how robust this effect is. Does it only emerge when people are answering relatively quickly and intuitively? Or would it emerge even if people took the time to reflect more carefully on the answers they were giving?

To address this question, we used a method that is already highly established and well-validated in a wide variety of different domains. All participants were given the Cognitive Reflection Test (CRT) (Frederick, 2005). In this test, participants receive a series of questions that are designed in such a way that a specific answer appears intuitively to be correct but can be seen, on reflection, to be mistaken. Existing studies have shown that participants who receive high CRT scores are more inclined to reject their initial intuitions on a variety of other tasks (Frederick, 2005), including on tasks that involve moral judgment (Hardman, 2012; Paxton, Ungar & Greene, 2011; Pinillos, Smith, Nair, Mun & Marchetto, 2011).

This effect becomes even stronger when the CRT is administered before the other tasks (Paxton et al., 2011; Pinillos et al., 2011). In that sort of design, the CRT itself becomes a manipulation. Those participants who receive high scores on the CRT have just experienced a task in which they saw that their own immediate intuitions were mistaken, and they therefore show an increased tendency to reflect on tasks that they receive subsequently within the same experimental session. So, for example, participants who receive high scores on the CRT and are then asked to make judgments of intentional action do not appear to show the usual impact of moral judgment (Pinillos et al., 2011).

Following the methods developed in this earlier research, we used the CRT to explore the role of cognitive reflection on judgments of innateness. If it is indeed the case that cognitive reflection serves to decrease or eliminate the impact of morality on these judgments, one would expect to find that participants with high CRT scores would show a smaller difference in judgments between the morally good and morally bad cases. Such an effect could arise if participants with high CRT scores were either (a) less inclined than other participants to attribute innateness in the morally good cases or (b) more inclined to attribute innateness in the morally bad cases. Whichever of these correlations obtained, one would expect it to be especially pronounced when the CRT was administered first.

Method

Participants. Two hundred twenty-two participants (47% female) filled out a questionnaire online using Amazon's Mechanical Turk.

Procedure. Each participant was assigned either to the *good* condition or to the *bad* condition. Participants in the good condition received both the 'abilities' vignette from Experiment 1 and the 'decent treatment' vignette from Experiment 2; participants in the bad condition received both the 'disabilities' vignette from Experiment 1 and the 'bad treatment' vignette from Experiment 2. The order of vignettes was counterbalanced, and after each vignette, participants received precisely the same question used in Experiments 1 and 2.

In addition, each participant received the Cognitive Reflection Test (CRT) (Frederick, 2005). Participants were assigned either to receive this test before reading the two vignettes or after reading the two vignettes.

Results

Each participant was given a score representing his or her mean response across the two items. These scores were then analyzed using a 2 (valence: good vs. bad) x 2 (order: CRT first vs. CRT second) ANOVA. Replicating Experiments 1 and 2, there was a main effect of valence such that participants were more inclined to regard the traits as innate in the good condition ($M = 4.6$, $SD = 1.4$) than in the bad condition ($M = 3.8$, $SD = 1.4$), $F\ (1, 218) = 15.2$, $p < .001$. There was no main effect of order, $F\ (1, 218) < .1$, $p = .84$, and no interaction effect, $F\ (1, 218) = .1$, $p = .79$.

To examine the relationship between participants' responses and their CRT scores, we then looked separately at the correlation between mean response and CRT score in each of the four cells of the design. Within the good condition, there was no significant correlation either when the CRT was administered first ($r = .19$, $p = .16$) or when it was administered second ($r = .23$, $p = .10$). Within the bad condition, there was no significant correlation when the CRT was administered first ($r = .13$, $p = .35$), and when the CRT was administered second, there was actually a significant correlation whereby participants with higher CRT scores regarded the trait as less innate ($r = -.33$, $p < 0.05$), meaning that participants with high CRT scores were even more inclined to offer the response one would expect if their innateness judgments were being impacted by moral judgments.

We then conducted an analysis using only those participants who received the highest possible CRT score ($n = 54$). Mean responses from these participants were analyzed using a 2 (valence: good vs. bad) x 2 (order: CRT first vs. CRT second) ANOVA. Even among these participants, there was a main effect of valence, with participants being more inclined to regard the traits as innate in the good condition ($M = 5.0$, $SD = 1.4$) than in the bad condition ($M = 4.0$, $SD = 1.4$), $F\ (1, 50) = 6.6$, $p < .05$. There was no main effect of order, $F\ (1, 50) = 2.5$, $p = .12$, and no interaction effect, $F\ (1, 50) = 1.6$, $p = .21$.

Discussion

This experiment used a well-established method to explore the impact of cognitive reflection on people's judgments of innateness. Overall, we did not find that the effect of moral judgment was decreased or eliminated among participants who were high in cognitive reflection. In fact, the results showed, if anything, a trend in the opposite direction, with participants who were high in cognitive reflection being even more inclined to deny innateness in the morally bad cases.

Experiment 4

Before turning to the judgments of trained scientists, we wanted to find a case in which the difference between scientists and ordinary folk would be predicted to go in the opposite direction. That is, we wanted a case in which people's ordinary folk judgments would not take a particular factor into account but in which scientists actually would regard that factor as relevant.

Accordingly, Experiment 4 turns to the distinction between capacities that are *learned* and those that are acquired in some other way. Existing analyses suggest that this distinction plays an important role in scientific uses of the concept of innateness (Carey, 2009; Cowie, 1999). However, there is reason to suspect that folk judgments of innateness will not show this same effect. In the theory of the folk concept of innateness developed by Griffiths and colleagues (Griffiths, 2002; Griffiths et al., 2009), it is claimed that people's ordinary judgments of innateness are based on certain features (fixity, typicality and teleology). Importantly, learning is not one of these features. Hence, the theory predicts that when all other factors are held constant, folk judgments of innateness will not be sensitive to the distinction between learned and non-learned capacities.

Griffiths and colleagues tested their theory by presenting participants with brief vignettes about the biology of bird species and showed that participants' judgments were sensitive to all three of the hypothesized features (fixity, typicality, and teleology). We now use that same method to check for a sensitivity to the distinction between capacities that are learned and those that are acquired in some other way.

Clearly, there is a general correlation whereby traits that are learned tend to be much lower in fixity than traits that are non-learned. To disentangle these two factors, we therefore present a pair of cases that are designed to be equal in fixity but to differ in the degree to which participants are likely to see the trait as learned.

Method

Participants. Sixty people (37% female, mean age 22, range 17-58) volunteered to fill out a questionnaire in the Yale University dining hall in exchange for $1.

Stimuli and procedure. Participants were randomly assigned to the 'learning condition,' the 'inference condition,' or the 'neuroscience condition.' Participants in the learning condition received the following vignette:

> Bird navigation is one of the most intensively studied aspects of animal behavior. Since the 1950s scientists have investigated in great detail the processes by which birds develop the ability to navigate.
>
> The Alder Flycatcher (Empidonax alnorum) is a migratory neo-tropical bird that breeds in southern Canada and the northern USA.
>
> Studies of the Alder Flycatcher show that, like many birds, they have the ability to use the sun as a 'celestial compass.' That is, they are able to combine information about the sun's position and the time of day, in order to determine direction of flight.
>
> Though this ability to navigate by the sun develops rapidly in fledgling Flycatcher, studies have shown that acquiring the ability requires approximately four hours visual experience in direct sunlight. This is required in order to learn the relationship between sun position and time of day, which is crucial to the operation of the bird's navigation system.
>
> As a matter of fact, virtually all Alder Flycatcher experience at least four hours of direct sunlight, and so virtually all members of the species develop the ability to navigate by the sun.

Participants in the inference condition received a vignette that was exactly the same except that the phrase 'learn the relationship between sun position and time of day' in the third paragraph was replaced with 'infer information about the relationship between sun position and time of day.' Participants in the neuroscience condition received a vignette that was exactly the same except that this phrase was replaced with 'activate a photosensitive region of the brain, called the suprachiasmic nucleus.'

All participants were then asked whether they agreed or disagreed with the statement: 'Navigation in the Alder Flycatcher is innate.' Answers were recorded on a scale from 1 ('disagree') to 7 ('agree').

Results and discussion

There was no significant difference between participants' responses in the learning condition ($M = 4.7$, $SD = 1.9$), the inference condition ($M = 4.3$, $SD = 2.0$) and the neuroscience condition ($M = 5.1$, $SD = 2.0$), $F\ (2, 57) = .73$, $p = .49$. In short, the results were exactly what one would predict on the Griffiths and colleagues (2009) theory.

Of course, the fact that we find no effect of learning in this specific study does not itself show that learning plays no role at all in folk judgments of innateness. It might well be possible to find an effect of learning in studies using other vignettes or using a design that differs in some other way. Future work could investigate these issues in greater detail.

Our concern here, however, is with a slightly different question: namely, with the ways in which scientific judgment diverges from folk judgment. We predict that when people are reasoning in a more scientific way, they will show a difference in innateness even on these exact cases. The question then is whether scientific reasoning does in fact lead to such a difference and, if so, whether their judgments in such cases are best explained by overwriting, by conceptual addition or by filtering.

Experiment 5

Existing scholarship has given us a good sense for the ways that scientists use the concept of innateness when they are making considered judgments in the context of their scientific research (Cowie, 1999; Griffiths, 2009; Kitcher, 1996; Mameli & Bateson, 2011; Samuels, 2002; Sober, 1999). What we see in Experiments 1-4 is that people's ordinary judgments seem to diverge in systematic respects from this scientific practice.

The question now is what explains this divergence. One possibility is that the divergence reflects a difference between two different groups of *people* (trained scientists vs. people who lack the relevant scientific training). Another possibility is that it reflects a difference between two different *modes of thought* (so that even people with no special training could show the 'scientific' pattern of judgments if they began thinking about the matter in the right way). The present experiment explores these two possibilities.

Participants were recruited from two very different populations. The sample of *folk* was composed of people who might be generally scientifically literate but who had no special training in the use of the concept of innateness. The sample of *researchers* was composed of people who were actively working as researchers in fields that used the concept of innateness. (Note that on this definition, a participant can count as 'folk' even if she has an in-depth understanding of certain kinds of research, as long as she has no special training in disciplines that use the concept of innateness.)

Within each of these groups, we compared the judgments participants made when they were focusing on *individual cases* to the judgments they made when they were focusing more on *general principles*. More specifically, we used the 'joint-separate' technique (Hsee, 1996). Some participants were assigned to receive just one condition from each of the pairs of cases in a between-subject design, while others were assigned to receive both versions and were asked explicitly whether there was any relevant difference between the two. This latter way of presenting the question tends to make participants think in a more principled way about which considerations are and are not relevant.

Note that the present manipulation makes it possible to address a different question from the one we explored above in Experiment 4. In that experiment, we asked

whether people who were generally more reflective would show a different pattern of innateness judgments. The answer appears to be no; participants who are asked to make judgments about individual cases tend to give the same pattern of answers regardless of how reflective they are. The present experiment does something further. Participants in the 'principled condition' are not simply encouraged in a general way to be more reflective; they are specifically encouraged to reflect about whether certain particular factors are relevant to innateness judgments. The experiment therefore makes it possible to determine whether people will arrive at a different pattern of judgments when they are encouraged to reflect in a principled way about the relevance of these factors.

Method

Participants. Participants logged on voluntarily to an online questionnaire study. Participants were recruited to the website using a variety of techniques. First, the popular science magazine *New Scientist* included a link to the study from its own site. Second, we posted information about the study to a number of scientific listservs, including listservs in psychology (Society for Judgment and Decision-Making, Society for Personality and Social Psychology, Society for Research in Child Development), philosophy (Society for Philosophy and Psychology) and linguistics (LINGUIST List). Third, we directly emailed all faculty working in genetics and molecular biology at the top 55 biology departments, as well as all faculty in evolutionary psychology and behavioral ecology listed on the website of the Human Behavior and Evolution Society (HBES).

In recruiting these participants, we collaborated with a number of other researchers. Each participant was randomly assigned to participate either in the present study or in a study conducted by these other researchers.

In total, 9472 people logged on to the site used for the present study, of which 6549 completed the entire questionnaire.

Procedure and stimuli. Each participant received three questions in random order: the Mother's Milk question (from Experiment 1), the Trait X question (from Experiment 2) and the Learning Question (from Experiment 3).

Each participant was assigned either to the case-based condition or to the principled condition. Participants in the case-based condition received each of the questions in precisely the same form used in the earlier experiments. Hence, for each question, each participant was randomly assigned to receive one or the other condition.

By contrast, participants in the principled condition received two versions of each vignette in a within-subject design. (For the Learning vignette, we used only the learning version and the neuroscience version, since the inference version was so conceptually similar to learning version.) Within each vignette type, the order of the two versions was counterbalanced.

Participants in the principled condition were told at the outset that the two versions differed only in a few words (which were underlined for easy identification) and that we specifically wanted to know whether they thought that this difference was relevant to whether the trait was innate. After reading each version, they were asked to rate their agreement with the statement about innateness for that version. Finally, after getting an innateness rating for each statement, they were given a few lines to 'explain why the difference between the two passages either was or was not relevant to the question of innateness.'

Finally, participants were asked whether they were working in philosophical or scientific research. Those who answered yes to this question were then asked to indicate an area of specialization from the options: 'psychology,' 'genetics,' 'linguistics,' 'biology' and 'other.' Each researcher was free to classify herself as falling into multiple categories.

Results

Out of a total of 6549 participants who completed the entire questionnaire, 1506 indicated that they were researchers. More specifically, there were 350 psychologists, 89 geneticists, 158 linguists, 435 biologists, 221 philosophers and 557 who indicated that they fell in some other category.

Means and standard deviations for each condition are displayed in Table 1. Analyses were conducted separately for each of the three questions.

Table 1. *Descriptive Statistics for Experiment 4*.

		Case-based		Principled	
		Folk	Researchers	Folk	Researchers
Mother's Milk	Ability	4.56 (2.27)	4.51 (2.15)	4.67 (2.35)	4.51 (2.22)
	Disability	3.95 (2.38)	3.72 (2.30)	4.36 (2.39)	4.26 (2.26)
Trait X	Decent	4.13 (2.32)	4.09 (2.29)	4.08 (2.38)	3.80 (2.32)
	Bad	3.40 (2.27)	3.65 (2.26)	3.94 (2.37)	3.70 (2.31)
Learning	Neuroscience	5.02 (2.18)	4.66 (2.20)	5.34 (2.12)	5.19 (2.08)
	Learning	4.55 (2.34)	4.78 (2.31)	4.16 (2.39)	4.13 (2.33)
	Inference	4.76 (2.22)	4.73 (2.19)	N/A	N/A

Mother's milk. In the case-based version, we again found that people gave higher ratings in the abilities condition ($M = 4.56$) than in the disabilities condition ($M = 3.90$), t (4297) = 9.45, $p < .001$. In the principled version, there was only a small difference between ratings for the abilities condition ($M = 4.62$) and the disabilities condition ($M = 4.33$), but because of the very large sample size, this small difference was statistically significant, t (2677) = 9.50, $p < .001$. (See Figure 1.)

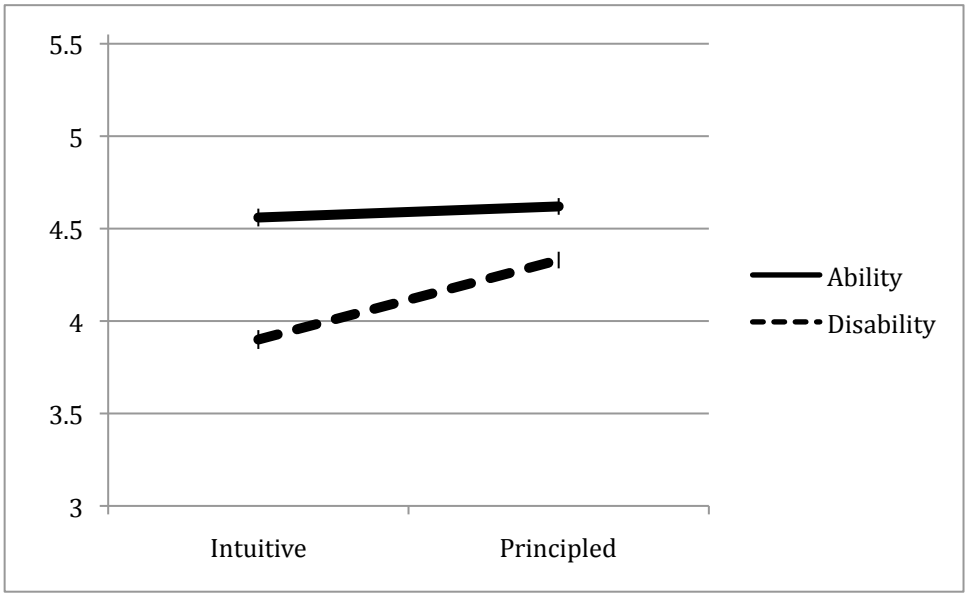

Figure 1. Mean innateness ratings for the mother's milk question. Error bars show SE mean.

To determine whether the effect for the case-based version was significantly greater than that for the principled version, we needed to compare a between-subject effect to a within-subject effect. We therefore computed effect sizes as r values for each design (appropriately correcting for dependencies in the within-subject comparison; Dunlap et al., 1996). Fisher's Z test on those values shows that the effect size for the intuitive version (r = .14) was significantly greater than the effect size for the principled version (r = .06), Z = 3.32, p < .001.

To further investigate this difference between intuitive and principled responses, we used separate t-tests to look at each of the pairwise comparisons. There was no significant difference between intuitive and principled responses within the abilities condition, $t(4824) = .90$, $p > .3$, but participants did give higher innateness ratings for the disability condition in the principled version, $t(4828) = 6.3$, $p < .001$.

We then compared the researchers to the non-researchers. For the case-based version, we conducted a 2 (researcher vs. non-researcher) x 2 (abilities vs. disabilities) ANOVA. There was no main effect of researcher, $F(1, 4020) = 2.7$, $p = .10$, and no significant interaction, $F(1, 4020) = 1.2$, $p = .27$. For the principled version, we conducted a 2x2 mixed-model ANOVA, with researcher vs. non-researcher as a between-subject factor and abilities vs. disabilities as a within-subject factor. Again, there was no main effect of researcher, $F(1, 2302) = 1.5$, $p = .23$, and no significant interaction, $F < 1$.

Trait X. In the case-based version, we again found that people gave higher ratings in the decent treatment condition ($M = 4.10$) than in the bad treatment condition ($M = 3.48$), $t(4306) = 8.90$, $p < .001$. In the principled version, this difference was small (3.92 to 3.79), but still significant, $t(2613) = 5.71$, $p < .001$. (See Figure 2.)

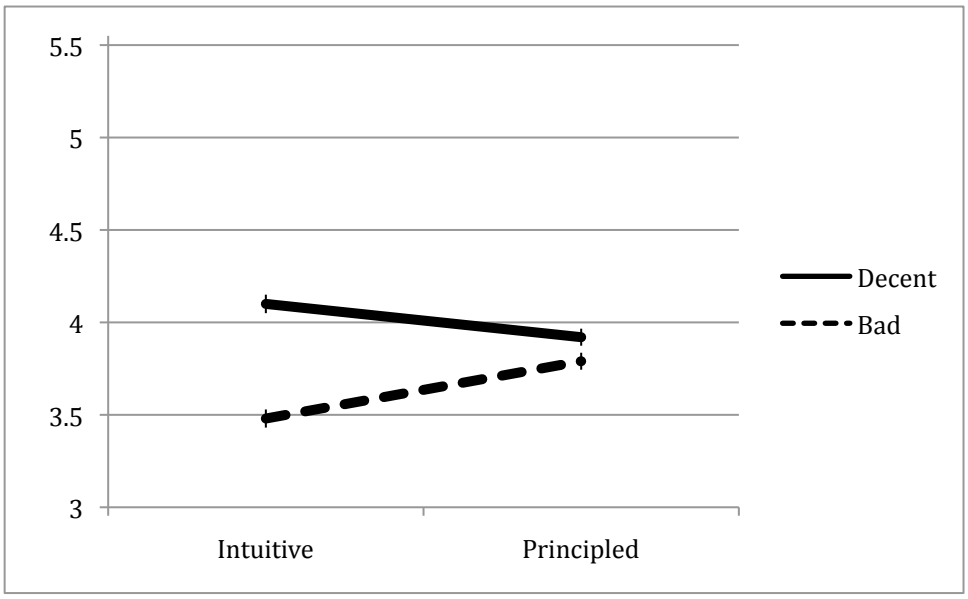

Figure 2. Mean innateness ratings for the Trait X question. Error bars show SE mean.

Here again, the effect size for the case-based version (r = .13) was significantly greater than that for the principled version (r = .03), Z = -4.4, p < .001. In the principled version, participants gave lower innateness ratings for the bad treatment condition, $t(4771) = 4.6$, $p < .001$, and higher innateness ratings for the decent treatment condition, $t(4761) = 2.7$, $p < .01$.

Researchers were compared to non-researchers using the same analyses described above for the Mother's Milk question. For the case-based version, there was no main effect of researcher, $F(1, 4019) = 1.4$, $p = .23$, and no significant interaction, $F(1, 4019) = 2.9$, $p = .09$. For the principled version, there was a main effect whereby researchers were less likely to regard the trait as innate in both conditions, $F(1, 2303) = 5.3$, $p < .05$, but, importantly, there was no significant interaction, $F < 1$.

Learning. In the case-based version, the difference between ratings for the learning condition ($M = 4.61$) and the inference condition ($M = 4.75$) was quite small, but still significant, $F(2, 4232) = 7.06$, $p = .001$. In the principled condition, there was a more substantial difference between the learning condition ($M = 4.2$) and the neuroscience condition ($M = 5.3$), $t(2599) = 27.54$, $p < .001$. (See Figure 3.)

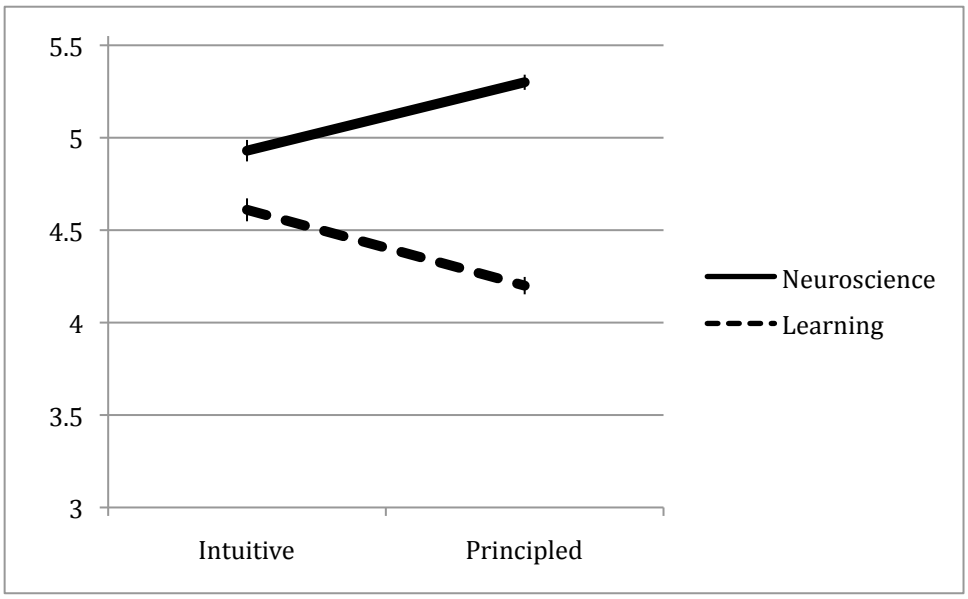

Figure 3. Mean innateness ratings for the learning question. Error bars show SE mean.

We compared the effect size in the case-based version for the contrast between the learning condition and the neuroscience condition (r = .07) to the principled version of that same contrast (r = .18). Fisher's Z test shows that the difference between these effect sizes is significant, Z = 4.07, p < .001. In the principled version, participants gave lower ratings for the learning condition, $t(4020) = 6.5$, $p < .001$, higher ratings for the neuroscience condition, $t(4009) = 4.9$, $p < .001$.

Researchers were compared to non-researchers using the method described above. For the case-based version, there was no main effect of researcher, $F < 1$, but there was a significant interaction, $F(2, 4018) = 4.3$, $p < .05$. An inspection of the means showed that the interaction arose because researchers were less inclined than non-researchers to regard the trait as innate in the neuroscience condition. For the principled version, there was no main effect of researcher, $F < 1$, and no significant interaction, $F(1, 2302) = 1.3$, $p = .25$.

Discussion

As predicted, when we looked at the case-based judgments of ordinary folks, we replicated the effects observed in Experiments 1-3. (People's intuitions were affected by moral considerations but not very much affected by learning.) Also as predicted, when we looked at the principled judgments of trained scientists, we found exactly what would be expected in light of existing scholarship on the scientific concept of innateness. (Scientists' reflective judgments were affected by learning but not very much by moral considerations.) The key aim of the study, however, was not merely to document the existence of this difference but rather to explain why exactly it arises. Is it fundamentally a difference between ordinary folks and trained scientists? Or is it a difference between case-based and principled thinking?

Here, the results were surprisingly unambiguous. Despite the very large sample size, we did not find systematic differences between ordinary folks and scientists. Instead, the effects seemed to be driven entirely by a difference between conditions. Participants assigned to the principled condition showed a substantially different pattern of judgments, and both ordinary folks and trained scientists showed this difference to roughly the same degree.

Admittedly, the participants in our 'folk' sample were recruited from the website of a popular science magazine, and they presumably had more background knowledge of science than one would expect to find in a more representative sample. Perhaps we would have obtained different results if we had looked at participants who were less scientifically literate (as in, e.g. Casler & Kelemen, 2008). Still, the results do seem to be pointing to something surprising about the role of scientific knowledge in people's responses. They indicate that whatever differences in knowledge there might be between a person who goes to popular science websites and a person who has a Ph.D. in cognitive science, these differences do not have a substantial impact on responses to the questions posed here.

Overall, then, the results suggest that the differences observed in these experiments are due not to a difference between different kinds of people (folk vs. scientists), nor to a difference between different psychological processes (intuition vs. reflection), but rather to a difference between different kinds of external situations. When people are placed in a situation that encourages them to think in a principled way about

the considerations that are influencing their judgments, they show a systematic tendency to shift toward that pattern of judgments described by existing scholarship on the scientific concept of innateness.

General Discussion

The concept of innateness plays an important role in contemporary cognitive science, but it also figures in people's ordinary folk understanding. The studies presented here were designed to explore the similarities and possible differences between folk and scientific usage. The results showed (a) that people's ordinary judgments about innateness were sensitive to moral considerations and (b) that, at least in certain cases, people's judgments showed surprisingly little sensitivity to learning-theoretic considerations. This pattern of responses was not affected by scientific training, but the results did vary depending on the way in which the question was framed. When the question was framed in a way that encouraged principled reflection about whether or not a given factor was relevant, the pattern was reversed, with people showing an effect for learning but very little effect of moral considerations.

In short, the pattern of judgments associated with the distinctively scientific use of the concept seemed to emerge not from intuitions about individual cases, but rather from principled reflection on which factors were relevant. At the level of judgments about individual cases, both scientists and non-scientists departed from the criteria found in the published scientific literature. However, at the level of more principled reflection, both scientists and non-scientists conformed to these criteria.

We now explore the implications of this finding on two different levels: first looking at the concept of innateness in particular, then treating the findings as a case study in a broader inquiry about the relationship between ordinary folk thought and scientific reasoning.

1. *The scientific concept of innateness*

The concept of innateness was not originally introduced by scientists. Long before the development of a systematic science of the mind, people were suggesting that certain

human capacities might be ‘innate’, and this notion played an important role in both Western and Eastern philosophy (Stich, 1975; Fung Yu-lan, 1953; Wong, 2012). Indeed, it has been suggested that the concept of innateness is a part of our folk biology – closely related to the notion that certain capacities are ‘built in’ or ‘in our nature’ (Linquist et al., 2011). But if the concept of innateness was not first introduced by scientists, a question arises as to how scientists have adapted this concept in such a way that it can now be used as part of scientific research. One possibility would be that scientists have abandoned the folk concept and replaced it with a distinctively scientific concept, either through what we earlier called overwriting or through conceptual addition. But our results provide evidence against both these hypotheses and in favor of what we call the filtering hypothesis.

If the overwriting hypothesis were correct, scientists should no longer have any vestige of the folk concept and should therefore show a distinctively scientific pattern of judgments even within the case-based condition. However, this result was not obtained. Instead, the finding was that even trained scientists showed a high impact of moral considerations and a low impact of learning when they were in the case-based condition. This finding provides evidence against the overwriting hypothesis.

If the conceptual addition hypothesis were correct, scientific training should offer people new conceptual resources that ordinary folks simply do not have. One should therefore expect scientists to differ from the folk in their judgments within the principled condition. But this in not what occurred. Instead, even participants who had no special training in cognitive science or related fields showed the same pattern evinced by scientists within the principled condition. These participants, too, said that moral considerations were not relevant but that learning was relevant. This indicates that whatever scientists are doing within the principled condition does not require some special knowledge or conceptual resources that other people lack.

In contrast, the results we obtained are exactly what one would expect on a filtering hypothesis. Within the case-based condition, both ordinary folks and trained scientists are influenced by moral considerations. However, when the experimental stimuli are designed in such a way that participants become explicitly aware that the question is about an influence of moral considerations, they filter out the result of their

usual intuition and instead conclude that this factor is not relevant. (Thus, participants in the principled condition were both less inclined to attribute innateness in morally good cases and more inclined to attribute innateness in morally bad cases.) Similarly, people's intuitions are sometimes insensitive to the distinction between learned and non-learned traits, but when they see explicitly that the question targets this distinction, they adjust their usual intuitions and conclude that the distinction is a relevant one.

A question now arises about why it is that people reject these specific case-based judgments when they are engaging in principled reflection. For the case of moral considerations, the answer may well be that people subscribe to a very general view according to which moral considerations cannot be relevant to the application of scientific concepts. Then, for the case of learning-theoretic considerations, it may be that participants are drawing on more domain-specific knowledge. Throughout contemporary cognitive science – and indeed throughout much of the history of the theoretical use of innateness – it is common practice to contrast the claim that a capacity is *innate* with the claim that this capacity is *learned* (e.g., Fodor, 1981; Cowie, 1999; Samuels, 2002). Assuming that our participants have been exposed, perhaps only in passing, to aspects of this scientific discussion, they may have a general understanding of the relationship between these two notions.

Overall, then, the present results point to an intriguing relationship between the scientific role of innateness and its role in people's folk understanding. It does not appear that scientists have replaced or supplemented the folk concept with a purely scientific one. Instead, it seems that scientists continue to use the folk concept but that, on reflection, they reject those aspects of the concept that they deem unhelpful in scientific research.

Finally, a question arises about how to reconcile the present results with existing scholarship on the scientific concept of innateness. This scholarship has not proceeded through experimental studies on individual scientists but rather by looking at the progress of naturally occurring scientific research programs as they unfold. As we have noted, the accounts developed within this tradition fit nicely with the results obtained in our principled condition but not with the results obtained in our case-based condition. How

can we now reconcile these case-based results with the findings from the scholarly literature?

One possible approach would be to start out with the assumption that the factors that have an impact in our case-based condition must also have an impact in actual scientific practice. If one starts with this assumption, the obvious conclusion would be that existing scholarship is simply mistaken and that scientific research on innateness actually shows an impact of moral considerations. Perhaps scientific researchers are systematically more inclined to regard traits as innate when they are produced by morally good environmental factors than when they are produced by morally bad environmental factors.

However, one might also reason in the opposite direction. If one starts with the assumption that existing scholarship is on the right track, one might conclude that the conditions that obtain in real scientific research are dissimilar to those found in our case-based condition. Certainly, there is ample additional reason to suppose that this conclusion is correct. Scientists are rarely asked to consider just one case in isolation; they are typically confronted with situations that encourage them to think systematically about a whole range of cases. Thus, the present results mesh nicely with a strain of thought within the study of science that emphasizes, not the ways in which scientists themselves differ from ordinary folks, but rather the ways in which the behavior of scientists is molded by characteristic features of their external situations (e.g., Kitcher, 1995; Mercier & Sperber, 2011).

2. *Morality and scientific judgment*

The impact of moral considerations observed in the present studies appears to be part of a broader pattern. Just as the present studies found an impact of moral considerations on people's use of the concept of innateness, earlier studies found an impact of moral considerations on people's use of the concepts of causation (Alicke, 2000; Hitchcock & Knobe, forthcoming); intentional action (Knobe, 2003; Nichols & Ulatowski, 2007); knowledge (Beebe & Buckwalter, forthcoming); freedom (Phillips & Knobe, 2009; Young & Phillips, 2011) and happiness (Phillips et al., 2011).

This phenomenon brings up an important general question. Experimental studies indicate that the use of a whole series of different concepts in ordinary thought can be influenced by moral considerations, but all of these concepts also appear in scientific research programs in which moral considerations are supposed to have no role. How is it possible for scientists to use these folk concepts in a less morally laden way?

One hypothesis would be that scientific reasoning shows an absence (or at least diminishment) of some of the psychological processes at work in ordinary thought. Ordinary thought appears to involve moral considerations, but one might suppose that when people are engaged in strictly scientific reasoning, they sometimes enter a distinct mode of thought in which moral considerations play no role. In other words, one might suppose that the reason why moral considerations do not influence the final judgments people reach in scientific reasoning is that they do not appear at any stage in the psychological processes leading up to those judgments.

The present studies do not provide a general test of this hypothesis, but they do allow us to explore it for the case of innateness in particular; and what the present results suggest is that the hypothesis is incorrect. It does not appear that people engage in scientific reasoning by entering a mode of thought in which moral considerations never play any role in the first place. Instead, the results suggest a more complex process. Even in scientific reasoning, people's case-based judgments continue to be influenced by moral considerations. However, when their attention is drawn to this influence, they tend to reject it and try to arrive at a pattern of judgments that sets moral issues aside and relies only on other considerations.

Further research could apply similar methods to other concepts (causation, intentional action, etc.). If such research arrives at converging results, we would have evidence for a more general claim about the role of moral considerations in scientific reasoning processes.

3. *Filtering*

Finally, at an even more general level, one can see the scientific use of the concept of innateness as just one example of a widespread phenomenon whereby scientists appropriate folk concepts for the purposes of systematic research. This

phenomenon can also be seen in the scientific use of semantic notions in linguistics (e.g. meaning, reference and truth), psychological concepts in the brain and behavioral sciences (e.g. preference, memory, emotion, goal, intention, and concept); and biological concepts in the life science (e.g. species, function and organism). In each of these cases, one can ask how it is possible for scientists to appropriate the relevant concept and use it in a research context.

Clearly, the answer in many cases will involve overwriting or conceptual addition. Scientists often develop new, more refined concepts that either replace or exist alongside the original folk versions. Thus the scientific notions of heat and temperature differ from anything in folk physics (Carey, 2009), the scientific notion of heritability (h^2) is deeply different from the folk notion of inheritance (Block, 1995), and many other concepts used in the sciences have departed in similar ways from their folk counterparts. The present studies suggest, however, that the appropriation of folk concepts can sometimes take a different form, which we have called *filtering*. In such cases, scientist's cognitive processes continue to rely on the folk concept, but when they see explicitly that the intuitions generated by the concept do not accord well with the aims of scientific research, they 'filter out' these intuitions and arrive at judgments that accord with more explicit principles. The research reported here explores this process for one particular case – the concept of innateness – but further work could look at other concepts (species, intention, memory, reference, etc.) and ask whether a similar process occurs for them as well. If such research were to arrive at converging results, we would have evidence for a more general claim about the role of filtering in scientific cognition.

Author Acknowledgments
We are grateful for comments and suggestions from Matthew Barker, Tim Bayne, Murray Clark, Brendan Dill, Paul Griffiths, Cecilia Heyes, Stefan Linquist, Edouard Machery, Ron Mallon, Eddy Nahmais, Jonathan Phillips, Nick Shea, Karola Stotz, Neil Van Leeuwen, Jonathan Weinberg and Dan Weiskopf, as well as audiences at Oxford University, Concordia University and Georgia State University.

Notes

[1] In phrasing the point this way, we rely on the assumption that these effects of value judgment on folk thinking reflect the structure of people's concepts, rather than the influence of some further factor that 'biases' or 'distorts' people's responses. (For a defense of this assumption, see, Knobe, 2010; but for opposing views, see Alicke et al., 2011; Levy, 2010.) This assumption will not be central to any of the key claims that follow, and we rely on it only for reasons of terminological convenience. Those who disagree can substitute for 'the folk concept of innateness' the phrase 'the psychological processes whereby ordinary folk arrive at judgments of innateness.'

References

Adams, F. & Steadman, A. (2004) Intentional action in ordinary language: Core concept or pragmatic understanding? *Analysis*, *64*, 173-81.

Alicke, M. (2000) Culpable control and the psychology of blame. *Psychological Bulletin*, *126*, 556-74.

Alicke, M., Rose, D. & Bloom, D. (2011). Causation, norm violation, and culpable control. *Journal of Philosophy*, *108*, 670-696.

Ariew, A. (1996). Innateness and canalization. *Philosophy of Science*, *63*, 19-27.

Beebe, J. R. & Buckwalter, W. (2010). The epistemic side-effect effect. *Mind & Language*, 25, 474-498.

Block, N. (1995). How heritability misleads about race. *Cognition*, *56*, 99-128.

Carey, S. (2009) *The origin of concepts*. Oxford: Oxford University Press.

Casler, K. & Kelemen, D. (2008). Developmental continuity in teleo-functional explanation: Reasoning about nature among Romanian Romani adults. *Journal of Cognition and Development*, *9*, 340-362.

Cowie, F (1999). *What's within? Nativism reconsidered*. Oxford: Oxford University Press.

Cushman, F., Knobe, J. & Sinnott-Armstrong, W. (2008). Moral appraisals affect doing/allowing judgments. *Cognition*, *108*, 353-80.

Ditto, P.H., Pizarro, D.A. & Tannenbaum, D. (2009). Motivated Moral Reasoning. In B. H. Ross (Series Ed.) & D. M. Bartels, C. W. Bauman, L. J. Skitka, & D. L. Medin (Eds.), *Psychology of Learning and Motivation, Vol. 50: Moral Judgment and Decision Making*. San Diego, CA: Academic Press, 307-338.

Douglas, H, (2009) *Science, policy, and the value-free ideal.* Pittsburgh, PA: University of Pittsburgh Press

Dunlap, W. P., Cortina, J. M., Vaslow, J. B., & Burke, M. J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods*, *1*, 170-177.

Fodor, J. 1981: The present status of the innateness controversy. In *Representations*. Cambridge, MA: MIT Press.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*, 25-42.

Fung Yu-Lan. (1953). *A history of Chinese philosophy*. 2 vols. E. J. Brill.

Gelman, S. (2003). *The essential child: Origins of essentialism in everyday life*. New York: Oxford University Press

Goldberg, R.F. & Thompson-Schill, S.L. (2009). Developmental 'roots' in mature biological knowledge. *Psychological Science*, *20*, 480-487.

Gopnik, A. & Wellman, H. (1992). Why the child's theory of mind really *is* a theory. *Mind & Language*, *7*, 145-171.

Griffiths, P. E. (2002). What is innateness? *The Monist*, *85*, 70–85.

Griffiths, P. (2009). The distinction between innate and acquired characteristics. *The Stanford Encyclopedia of Philosophy (Fall 2009 Edition)*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2009/entries/innate-acquired/>.

Griffiths, P. E. Machery & S. Linquist (2009). The vernacular concept of innateness. *Mind and Language 24*, 605-630.

Hardman, D. (2012). Moral dilemmas: Who makes utilitarian choices? Unpublished manuscript, London Metropolitan University.

Hitchcock, C. & Knobe, J. (2009) Cause and norm. *Journal of Philosophy*, 106, 587-612.

Hsee, C. K. (1996). The evaluability hypothesis: An explanation for preference reversals between joint and separate evaluations of alternatives. *Organizational Behavior and Human Decision Processes*, *67*, 247-257.

Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Straus & Giroux.

Keil, F. C. (1992). *Concepts, kinds, and cognitive development*, MIT Press: Cambridge, MA.

Kelemen, D., Rottman, J., & Seston, R. (2012, in press). Professional physical scientists display tenacious teleological tendencies: Purpose-based reasoning as a cognitive default. *Journal of Experimental Psychology: General*.

Khalidi, M. A. (2002). Nature and nurture in cognition. *British Journal for the Philosophy of Science*, *53*, 251-272.

Khalidi, M. A. (2007). Innate cognitive capacities. *Mind & Language*, *22*, 92-115.

Kitcher, P. (1995). *The advancement of science: Science without legend, objectivity without illusions*. New York: Oxford University Press.

Kitcher, P. (1996). *The lives to come*. New York: Simon & Schuster.

Knobe, J. (2003). Intentional action and side effects in ordinary language. *Analysis*, 63,190-3.

Knobe, J. (2010). Person as scientist, person as moralist. *Behavioral and Brain Sciences*, 33, 315-329.

Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review*, 96, 674-689.

Kuhn, D., Schauble, L. & Garcia-Mila, M. (1992). Cross-domain development of scientific reasoning, *Cognition and Instruction*, *9*, 285-327.

Lehrman, D. S. (1970). Semantic and conceptual issues in the nature-nurture problem. *Development & Evolution of Behaviour*, D. S. Lehrman (ed.), San Francisco: W. H. Freeman and Co: 17–52.

Leslie, A., Knobe, J. & Cohen, A. (2006). Acting intentionally and the side-effect effect: 'Theory of mind' and moral judgment. *Psychological Science*, *17*, 421-7.

Levy, N. (2010). Scientists and the folk have the same concepts. *Behavioral and Brain Sciences, 33*, 344-344.

Linquist, S., Machery, E., Griffiths, P. E. & Stotz, K.. (2011). Exploring the folkbiological conception of human nature. *Philosophical Transactions of the Royal Society B*, *366*, 444-454.

Lombrozo, T. , Kelemen, D. & Zaitchik, D (2007). Inferring design: Evidence of a preference for teleological explanations in patients with alzheimer's disease. *Psychological Science, 18,* 999-1006.

Machery, E. (2006). The folk concept of intentional action: Philosophical and experimental issues. *Mind & Language*, *23*, 165–189.

Mallon, R. and Weinberg, J. (2006) Innateness as closed-process invariance. *Philosophy of Science*. *73*, 323-344.

Mameli, M. & Bateson, P. (2006). Innateness and the sciences, *Biology and Philosophy*, 21, 155-188.

Mameli, M. & Bateson P. (2011). An evaluation of the concept of innateness. *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences*, *366*, 436-443.

McCloskey, M. (1983). Naïve theories of motion. In D. Gentner and A. L. Stevens (Eds.), *Mental Models* (pp. 299-324). Hillsdale, NJ: Erlbaum.

Menzies, P. (2010). Norms, causes and alternative possibilities. *Behavioral and Brain Sciences, 33*, 346-347.

Mercier, H. & Sperber, D. (2011) Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, *34*, 57-74.

Nichols, S. & Ulatowski, J. (2007). Intuitions and individual differences: The Knobe effect revisited. *Mind & Language*, *22*, 346–365.

Paxton, J.M., Ungar, L. & Greene, J.D. (2011). Reflection and reasoning in moral judgment. *Cognitive Science*, *36*, 163-177.

Phillips, J. & Knobe, J. (2009). Moral judgments and intuitions about freedom. *Psychological Inquiry*, 20, 30-36.

Phillips, J. Misenheimer, L. & Knobe, J. (2011). The ordinary concept of happiness (and others like it). *Emotion Review*, *71*, 929-937.

Pinillos, Á., Smith, N., Nair, G. S., Mun, C. & Marchetto, P. (2011). Philosophy's new challenge: Experiments and intentional action. *Mind & Language*, *26*, 115-139.

Samuels R. (2002). Nativism in cognitive science. *Mind & Language*, *17*, 233-265.

Samuels, R. (2004). Innateness and cognitive science. *Trends in Cognitive Sciences. 8*, 136-141.

Scholl, B. J. & Leslie, A. M. (1999), Modularity, development and 'theory of mind'. *Mind & Language*, *14*, 131-153.

Shtulman, A. (2006). Qualitative differences between naïve and scientific theories of evolution. *Cognitive Psychology*, *52*, 170-194.

Shtulman, A. & Valcarcel, J. (2012). Scientific knowledge suppresses but does not supplant earlier intuitions. *Cognition,* 124, 209-215.

Slotta, J. D. & Chi, M. T. H. (2006). Helping students understand challenging topics in science through ontology training. *Cognition and Instruction*, *24*, 261-289.

Sober, E. (1999). Innate knowledge. In *The Routledge Encyclopedia of Philosophy*,Vol. *4*, 794-797. Routledge.

Sripada, C. & Konrath, S. (2011). Telling more than we can know about intentional action. *Mind & Language*, *26*, 353-380.

Stich, S. P. (1975). The idea of innateness. *Innate Ideas*, S. P. Stich. Los Angeles: University of California Press.

Sytsma, J., J. Livengood, and D. Rose (forthcoming). Two types of typicality: Rethinking the role of statistical typicality in ordinary causal attributions. *Studies in History and Philosophy of Science.*

Uttich, K. & Lombrozo, T. (2010). Norms inform mental state ascriptions: A rational explanation for the side-effect effect. *Cognition*, *116*, 87-100.

Weisberg, D. S, Keil, F.C., Goodstein, J., Rawson, E. & Gray, J. R. (2008). The seductive allure of neuroscience explanations. *Journal of Cognitive Neuroscience*. *20*, 470-7.

Wong, D. (2012). How early Confucian philosophy helps us to think about the nature and nurture of moral development. Philomathia Lectures on Human Values. Chinese University of Hong Kong.

Young, L. & Phillips, J. (2011). The paradox of moral focus. *Cognition*, *119*, 166-178.

Young, L., Cushman, F., Adolphs, R., Tranel, D. & Hauser, M. (2006). Does emotion mediate the effect of an action's moral status on its intentional status? Neuropsychological evidence. *Journal of Cognition and Culture*, *6*, 291-304.

Zalla, T. & Leboyer, M. (2011). Judgment of intentionality and moral evaluation in individuals with high functioning autism. *Review of Philosophy and Psychology* 2, 681-698.