

## 35

# The Concept of Innateness as an Object of Empirical Enquiry

RICHARD SAMUELS

## 35.1 Introduction

The concept of innateness has led a busy life. Within the sciences, it has historically exerted an influence in many regions of biology; and though its explicit role in the life sciences has waned in recent decades, it continues to play a significant role in cognitive science – especially, developmental psychology and linguistics. Yet, the concept of innateness is quite unlike many that figure in science. Whereas the concepts of a Higgs boson or lateral geniculate nucleus, for example, are products of science and make little sense independently of this context, the concept of innateness has led a life outside the lab. Discussions of innateness appear in works of both Western philosophy (Cowie 1999) and Chinese philosophy (Fung 1953; Wong 2012) that long predate modern science. Moreover, the concept appears to have a place in our ordinary folk understanding as well. Even if many people rarely come across the actual English word “innate,” they can easily understand the claim that certain capacities are “just built into us,” a product of nature, or “in our genes.”

The role of the concept of innateness – especially in the sciences – has also been contentious. According to the standard criticism, widely endorsed by philosophers and theoretically oriented biologists, the concept of innateness is unfit for scientific purposes because it encodes a prescientific conception of development that conflates several distinct biological issues and thereby leads researchers to commit fallacies of ambiguity (Bateson 1991; Griffiths 2002). Furthermore, the most vociferous recent advocates of this standard criticism maintain that it is bolstered by empirical research on innateness judgments (Griffiths, Machery, and Linquist 2009).

This chapter has a pair of aims. The first is to provide an overview of some recent efforts to empirically study the innateness concept, both as deployed in folk contexts and among scientists. The second aim is to consider whether this research really bolsters the standard criticism. In Section 35.2, I describe research by Paul Griffiths and his collaborators, which seeks to assess

whether the folk concept of innateness is a manifestation of our folk biology. In Section 35.3, I consider whether, as Griffiths *et al.* maintain, this research bolsters the standard criticism. In Section 35.4, I review further research, largely due to Josh Knobe and Richard Samuels, on folk innateness judgments; and in Section 35.5 I describe how this research was extended in order to explore the issue of whether scientists' innateness judgments rely on a distinctly scientific innateness concept, or whether they merely redeploy the folk concept. Finally, in Section 35.6, I conclude with some very brief comments on the implications of this research.

## 35.2 Folk Innateness Judgments: The Innateness Concept as a Manifestation of Our Folk Biology

Philosophers have long sought to analyze the concept of innateness (Stich 1975); and over the past few decades a number of explicit definitions have been proposed. (For surveys see Samuels 2004; and Mameli and Bateson 2011). Experimental studies of the innateness concept are quite different in emphasis. Rather than seeking to define “innate,” the goal has been to describe the sort of mental processes and representations that underlie our thought about innateness. In this section, I set out one intriguing proposal, initially suggested by Paul Griffiths (2002), and later explored by Griffiths in collaboration with Edouard Machery, Stephan Linquist, and Karola Stotz. According to this proposal, which they call the *three-feature hypothesis*, the concept of innateness is a manifestation of certain core aspects of our folk biology (Griffiths 2002; Griffiths, Machery, and Linquist 2009; Linquist *et al.* 2011).

### 35.2.1 Background: Folk Essentialism

Let me start by laying some background in place. Over the past four decades, naïve thought about the biological world has been a focus of intensive study among cognitive psychologists, anthropologists, and developmentalists. Although much controversy remains (see, e.g., Sloutsky, Kloos, and Fisher 2007), there is fairly widespread consensus concerning certain core aspects of our folk biology – aspects that Griffiths and his collaborators rely on in developing their hypothesis. Crucially, there is considerable agreement that quotidian thought about the biological realm is a manifestation of *folk essentialism*. When ordinary people draw inferences or make judgments about biological phenomena, they tend to suppose that organisms possess a hidden *causal essence* or *inner nature* — a property or set of properties that is possessed by the organism throughout its life, and which both defines its kind membership and causes it to possess kind-typical properties (Medin and Ortony 1989; Atran 1990; Gelman 2003). On this view, for example, people tacitly suppose that what makes an organism a *tiger* with typical tigerish characteristics is that throughout its life it possesses a hidden species essence that causes the organism to exhibit such species typical properties as being four-legged, furry, and striped. Furthermore, this bias appears not to be an artifact of local cultural mores. Rather, what evidence there is suggests that folk essentialism is pan-cultural, and that it emerges reliably, and quite early in development (Atran 1998).

Although folk essentialism has been invoked to explain a broad array of psychological phenomena, for present purposes the most relevant is a tendency, shared by adults and children alike, to suppose that species typical traits are insensitive to environmental variation. For example, when asked to imagine a cow that has been raised by pigs, adults routinely assume that the cow will display normal bovine traits, such as mooing instead of oinking (Atran *et al.* 2001). Folk essentialism readily explains this fact because if people assume that an organism will possess the same species essence

throughout its life, then an organism's species typical traits – which are a product of this essence – should be expected to develop across an exceedingly broad range of environments.

### 35.2.2 *The Three-Feature Hypothesis*

We are now in a position to set out the three-feature hypothesis. It can be divided into two main parts. The first is that folk judgments about innateness reflect our ordinary tendency to think in essentialist terms about biological kinds (Griffiths 2002; Griffiths, Machery, and Linquist 2009). Specifically, Griffiths and his collaborators propose the following:

*Essentialist Thesis:* Naïve subjects judge a trait as innate when it is categorized as a product of the organism's species essence.

But on what basis do people work out whether a trait is an expression of a species essence and, hence, innate? Here's where the second part of the hypothesis comes in. According to Griffiths and his collaborators:

*Categorization Thesis:* Naïve subjects categorize a trait as a product of the organism's species essence largely on the basis of the extent to which it exhibits three features: fixity, species typicality, and teleology (or functionality).

Roughly, a trait is *fixed* to the extent that its possession is unaffected by environmental variation; a trait is *species typical* to the extent that it is possessed by all members of the species that are not abnormal; and a trait is *teleological* to the extent that it contributes to the well-being of the organism which possesses it. Other factors may, Griffiths *et al.* suppose, exert an influence; but these three are central. Furthermore, and more stringently, they hypothesize that evidence of fixity, typicality and teleology are *independent* of each other, and consequently, they predict that these three features will contribute *additively* to judgments about whether some trait is innate.

The above earlier proposal requires a bit more unpacking. In particular, it is important to appreciate how the Essentialist and Categorization theses are connected – why the Essentialist Thesis should lead Griffiths *et al.* to suppose that innateness judgments are largely determined by fixity, typicality, and teleology. In light of our earlier discussion of folk essentialism, it should be pretty obvious why typicality and fixity are relevant. Evidence that a trait is species typical should increase the probability that it is judged to be an expression of a species essence, hence innate, because, by assumption, species typical traits are products of a species essence. Similarly, evidence that a trait is fixed should raise the probability that it is judged innate because, by hypothesis, essences are invariant over the lifespan of the organism, and the traits they produce are insensitive to environmental influence.

It is rather less obvious, however, why evidence of functionality should raise the probability that a trait is judged innate. As Griffiths *et al.* clearly recognize, this assumption does not flow directly from the hypothesis that we are folk essentialists. As a consequence, they cite various auxiliary considerations in order to bolster the claim that teleology is relevant. For example, they note the cross-cultural tendency for people to invoke teleology in explaining why animals and plants possess traits (Atran 1995); and (quite unrelatedly) they note that some scientists and philosophers have suggested that “innate” should be analyzed to mean “designed by natural selection.” These considerations are, however, clearly auxiliary to the core idea that innate traits are an expression of an organism's species essence; and as such, the status of teleology seems quite unlike that of fixity and typicality. I return to this point later on, but first we need to consider the empirical support for the three-feature hypothesis.

### 35.2.3 *Experimental Study*

Suppose the three-feature hypothesis is correct. Then we should expect the following:

Prediction 1: When a trait is associated with any of the three features – fixity, typicality and teleology – participants will be more likely to judge the trait innate.

Prediction 2: All three features will contribute *independently* to participants' judgments about the trait's innateness.

To test these predictions, Griffiths and his collaborators ran a series of studies in which they presented vignettes describing the song of different bird species (Griffiths, Machery, and Linquist 2009; Linquist *et al.* 2011). The vignettes systematically varied whether a trait was species typical, whether its development was dependent on the environment and whether it was functional. More specifically, each probe was organized in the following manner:

- First, there is an initial paragraph about research on birdsong, designed to convince participants that there is a wealth of well-established scientific knowledge about birdsong.
- The next paragraph begins with one or two sentences, designed to convince participants that the animal is real, naming a specific bird and providing some neutral information about it.
- The remainder of the second paragraph states whether the song of the male of this species is fixed, typical, teleological, or their opposites, using one of each of these pairs of statements:

Fixed/Plastic

0. Studies on \_\_\_\_\_ show that the song an adult male produces depends on which songs they hear when they are young.
1. Studies on \_\_\_\_\_ show that the song an adult male produces does not depend on which songs they hear when they are young.

Typical/Typical

0. Studies also show that different males in this species sing different songs.
1. Studies also show that all males of this species sing the same song.

Teleology/Teleology

0. Close observations of these birds reveal that the males' song is not used to attract mates or to defend territories. Scientists therefore agree that this feature of the bird has no real function, like the appendix in humans.
  1. Close observations of these birds reveal that the males' song attracts mates and helps to defend their territory. Scientists therefore agree that this feature of the bird has a real function, like the heart in humans.
- Finally, the probe concludes with a question of the following form:  
On a 7-point scale, 1 meaning strongly disagree and 7 meaning strongly agree, how would you respond to the following statement?  
"The song of the male \_\_\_\_\_ is innate."

To illustrate, here is the probe describing a species of bird in which birdsong is not-typical but is fixed and has a function:

Birdsong is one of the most intensively studied aspects of animal behaviour. Since the 1950s scientists have used recordings and sound spectograms to uncover the structure and function of birdsong. Neuroscientists have investigated in great detail the areas of the brain that allow birds to develop and

produce their songs. Other scientists have done ecological fieldwork to study what role song plays in the lives of different birds.

The Alder Flycatcher (*Empidonax alnorum*) is a migratory neo-tropical bird which breeds in southern Canada and the northern USA. Studies on the Alder Flycatcher show that the song an adult male produces does not depend on which songs they hear when they are young. Studies also show that different males in this species sing different songs. Furthermore, close observations of these birds reveal that the males' song attracts mates and helps to defend their territory. Scientists therefore agree that the bird's song has a real function, like the heart in humans.

On a 7-point scale, 1 meaning strongly disagree and 7 meaning strongly agree, how would you respond to the following statement?

"The song of the male Alder Flycatcher is innate."

Griffiths and his collaborators used vignettes of this sort in three different studies, two of which have a between-subjects design, another within-subjects. What they consistently found is that (1) typicality and fixity are both significant predictors of folk judgments about innateness; (2) fixity explains more of the variance than typicality; and (3) the influence of these two factors is additive. These results comport well with the three-feature hypothesis. In contrast, the data on teleology is far less clear, and fit far less well. At best, teleology appears to be a marginally significant predictor of innateness judgments – explaining only a very small part of the variance. But in two of their experiments – both of which had between-subject designs – teleology had no significant effect at all on judgments of innateness.

#### 35.2.4 Follow-up Study

The three-feature hypothesis predicts that teleology should exert a significant influence on folk innateness judgments. Why then did it exert so little influence, and what implications does this have for the three-feature hypothesis? Clearly, the data provide little reason to reject the Essentialist Thesis. But it may appear to undermine the Categorization Thesis since only two of the three proposed features make a significant contribution to folk innateness judgments. But *why* is it that teleology exerted so little influence? This is a question that Griffiths and collaborators sought to address in a follow-up study (Linguist *et al.* 2011).

Perhaps the most obvious explanation for why teleology exerts so little influence is that, contrary to what Griffiths *et al.* assume, the Categorization Thesis is false: Our folk biology does not construe essence-produced traits as teleological. And, as a matter of fact, there is little direct evidence that subjects do view essence-produced traits in this way. This is why, as noted earlier, when Griffiths *et al.* argue for the inclusion of teleology, they resort to considerations that are more-or-less independent of the Essentialist Thesis. But if this is true – if there is little reason to suppose that subjects think of essence-linked traits as functional – then Griffiths *et al.*'s data actually fit *better* with the Essentialist Thesis than would have been the case had teleology turned out to be an important factor. In short: on the present view, the evidence against the Categorization Thesis actually supports the Essentialist Thesis.

There is a second intriguing possibility, however, which Linguist *et al.* (2011) explore. Rather than rejecting the Categorization Thesis, they suggest that our folk biology *does* treat essence-produced traits as teleological, but that teleology fails to influence verbal assertions involving the English word "innate" because – for whatever reason – "innate" is not especially effective at "tapping into" our underlying folk biological theory. If this were true, they suggest, one might expect other English expressions to do a better job at satisfying the Categorization Thesis; and as matter of fact, this appears to be the case – at least among Anglophone undergraduates. Specifically, in their 2011 study, Linguist *et al.* posed tasks very similar to those outlined earlier, except that they used the expression "in its DNA" instead of "innate." This manipulation was

guided by the hunch that the relatively colloquial expression “in its DNA” would – at least for naïve undergraduate participants – do a better job of tapping into our folk biology than the somewhat antiquated word “innate.” What they found was that for judgments concerning whether an organism’s trait is “in its DNA,” the three-features – fixity, typicality, and teleology – explain far more of the variance than they do for innateness judgments: 46% as opposed to 16%. Moreover, they found that teleology alone is a good predictor of judgments about whether a trait is in the DNA, and indeed a *better* predictor than typicality is.

Although the findings outlined in the previous paragraph are consistent with the three-feature hypothesis, they also raise a range of issues that require further clarification and exploration. One such issue concerns why we should accept the assumption that “in its DNA” is eliciting judgments that depend on the very same cognitive structures as those elicited by “innate.” Minimally, in its present form, the three-feature hypothesis leaves it a mystery why an inversion in the significance of teleological considerations should occur, if “innate” and “in its DNA” both rely on the same essentialist folk biology. Furthermore, there is an obvious alternative hypothesis that merits consideration. Given the recent and purely scientific provenance of talk about DNA, it may be that “in its DNA” elicits judgments that rely on a distinct but related set of semi-scientific concepts. Among other things, such a proposal would explain the heavier weighting of teleological considerations, if (as I suspect) many people associate DNA with the process of evolution by natural selection – a commonly made connection is popular scientific works, such as Dawkins’ *The Selfish Gene*.

### 35.3 Does Empirical Research on the Three-Feature Hypothesis Bolster the Standard Criticism?

Although one major goal of Griffiths *et al.*’s research is simply to understand the psychological basis of innateness judgments, they also maintain that it has implications for longstanding debates over the scientific legitimacy of the innateness concept. Specifically, they maintain that:

The data from our ... empirical studies bolsters the standard scientific criticism of the concept, which is that it conflates a number of different ideas and leads to fallacies of ambiguity (Griffiths, Machery, and Linquist 2009, 624).

In what follows, I set out this criticism in more detail, and argue that the extant research fails to support this contention.

#### 35.3.1 *Innateness as a “Tonkish” Concept*

In a well-known paper on logical connectives, Arthur Prior discussed a connective, TONK, which everyone would consider *defective* because it permits, via repeated applications of its introduction and elimination rules, the inferring of anything from anything else (Prior 1967). Griffiths and his collaborators, of course, never suggest that innateness is nearly so permissive. But they do suggest that it is “tonkish” in that it licenses illicit patterns of inference from one claim that we have reason to endorse, to others that are entirely unwarranted (Machery, 2014). Specifically, they contend that the innateness concept permits one to illegitimately infer from the trait’s possession of one of the three features, to its possession of the others.

Suppose, for example, that one has evidence that an organism’s fur color is species-typical. Then since the influence of typicality on innateness judgments is independent, one might make the following “introduction” inference:

Trait x is typical  
Trait x is innate

More generally, because the three features exert an independent influence

... people need not know whether a trait is fixed or has a function to decide whether it is innate on the basis of evidence about typicality (and vice versa). Thus, if they are told that a trait is species-typical, people may well infer that it is innate.

But presumably we also make “elimination” inferences, from a trait’s being innate to its possession of other properties – for example,

Trait x is innate  
Trait x has a function

But if this is so, then the concept of innateness would appear to license inferences from justifiable claims –in our example, the trait’s species-typicality – to quite different, and entirely unsupported conclusions (e.g., that the trait is functional). As Griffiths *et al.* put it:

Having judged that [a trait] is innate, people are likely to infer that it is fixed—that its development does not depend on its environment. Or if they are told that a trait has a function, people may infer on that basis alone that it is innate. Having judged that it is innate, they are likely to conclude that it is species-typical (and so on).

In which case, it might seem that the innateness concept is tonkish and, hence, defective. Furthermore, if innate is defective in this fashion, then it may seem reasonable to conclude that it ought not to be deployed in the sciences, where the avoidance of equivocation and intellectual confusion are presumably quite desirable.

### 35.3.2 *Why the Extant Empirical Research Fails to Bolster the Standard Criticism*

The research outlined earlier fails to bolster the Standard Criticism. First, in order to sustain the claim that the innateness concept is tonkish, one needs evidence concerning both introduction inferences (e.g., from fixity to innateness) *and* elimination inferences (e.g., from innateness to typicality). It is only if both exhibit the right inferential profile that the concept will behave in a tonkish manner. But as a matter of fact, the Griffiths *et al.* studies say nothing whatsoever about elimination inferences. Rather, they wholly concern introduction inferences. And what the evidence at most suggests is that innateness judgments are made on the basis of multiple dissociable factors. The same is true, however, of many concepts – including many that figure in scientific contexts. For example, scientists judge a substance to be a *mineral* on the basis of multiple dissociable factors – for example, whether it is naturally occurring, stable at room temperature, and abiogenic. But it would surely be implausible to conclude from this alone that the mineral concept should be expunged from science.

Second, it’s unclear how Griffiths *et al.*’s studies could provide very much support for the standard criticism, since it’s far from clear that they tell us much about innateness judgments in science. The point is an obvious one. The research conducted by Griffiths and his collaborators focuses on *folk* applications of the term “innate.” But such research will not do much to bolster the standard criticism, unless we have reason to suppose that innateness judgments in science deploy the very same conceptual resources. Yet, the studies outlined earlier are silent on this matter. Moreover, we ought not to assume that the same concept is operative merely because the

same word, “innate,” is being used in both contexts. After all, there are a great many examples of words that express quite different concepts in folk and scientific contexts. “Heat,” “weight” and “velocity” come readily to mind (Carey 1992). In which case, even if the earlier studies established the tonkishness of the folk innateness concept – which they do not – this alone would do little to support the contention that scientific usage is similarly problematic.

One obvious question, then, is whether scientists deploy the same concept of innateness as the folk? In Section 35.5, I focus on this issue. But first I need to describe another strand of research on folk innateness judgments – this time from work by Josh Knobe and Richard Samuels (Knobe and Samuels 2013).

## 35.4 More on Folk Innateness Judgments

In their work on folk innateness judgments Knobe and Samuels focused primarily on two questions:

- To what extent do learning theoretic considerations influence folk innateness judgments?
- Are folk innateness judgments influenced by value judgments?

Both questions were motivated, in large measure, by what they saw as central aspects of the way in which cognitive scientists use the concept of innateness. Specifically, on the basis of an assessment of published research on innateness hypotheses, they argued that learning theoretic considerations appear very important to whether cognitive scientists count a trait as innate, but that considerations of value appear, on the face of it, to exert little or no influence. As such, Knobe and Samuels’ work on folk judgments was a preliminary to considering whether scientists in the relevant fields deploy a folk concept of innateness or some distinctly scientific one.

### 35.4.1 Learning and Folk Innateness Judgments

It is widely supposed in cognitive science and allied disciplines that there is an important connection between innateness and learning. Indeed some very influential cognitive scientists go so far as to maintain that the two concepts are inter-defined – that innate traits just are those that are not the output of learning processes (Carey 2010). But even if this claim about the definition of innateness is untrue, it is still the case that in published research it is invariably assumed that a trait will not count as innate, if an organism acquires it by perceiving its environment and engaging in straightforward learning. Within cognitive science, then, it appears to be no more than a *banal truism* that learned traits are not innate.

Should we expect folk innateness judgments to conform to this banal truism? If the three-feature hypothesis is correct, then we should not. More precisely, the hypothesis predicts that when all three features are held constant, folk innateness judgments should not be sensitive to the distinction between learned and non-learned capacities.

In order to test this prediction, Knobe and Samuels presented vignettes to 60 Yale students that were based on the “bird” vignettes used by Griffiths and his collaborators. But rather than varying the fixity, typicality or functionality of the trait, participants were instead randomly assigned to a “learning condition” in which the trait was said to have been learned or to a “neuroscience condition” in which the trait was described as the product of a “brute casual” process. 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 neuroscience condition received a vignette that was exactly the same except that this phrase “learn the relationship between sun position and time of day” in the third paragraph was replaced with “activate a photosensitive region of the brain, called the suprachiasmatic 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”).

The results were exactly what one would predict if the three-feature theory were correct. In other words, there was no significant difference between participants’ responses in the learning condition ( $M = 4.7$ ,  $SD = 1.9$ ), and the neuroscience condition ( $M = 5.1$ ,  $SD = 2.0$ ). But if this is so – if people exhibit insensitivity to learning-theoretic considerations – *why* does scientific research that deploys the concept of innateness invariably conform to the truism that learned traits are not innate? We return to this issue in Section 35.5.

### 35.4.2 Value and Folk Innateness Judgments

As mentioned earlier, considerations of value appear to exert little or no influence in which traits are claimed by cognitive scientists to be innate. Yet over the past decade or so, there has been a steady accumulation of research suggesting that value judgments exert a surprisingly large influence on ordinary folk thought about apparently straightforwardly factual matters. 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 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, and Tannenbaum 2009; Knobe 2003; Nichols and Ulatowski 2007, Cova this volume; but see Machery 2006). Related effects have been observed for people’s use of numerous other concepts, including the concepts of causation (Alicke 2000; Hitchcock and Knobe 2009; Livengood and Rose this volume), knowledge (Beebe and Buckwalter 2010; Beebe this volume), freedom (Phillips and Knobe 2009; Young and Phillips 2011; Chan *et al.* this volume) and the distinction between doing and allowing (Cushman, Knobe, and 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).

Against this background, Knobe and Samuels sought to determine whether value judgments exert a similar influence on folk innateness judgments. In one experiment, participants were assigned either to an “abilities condition” or to a “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").

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 ( $M = 3.3$ ,  $SD = 1.7$ ).

Although these results are suggestive of the claim that innateness judgments are influenced by value judgments, taken in isolation, this result is clearly consistent with alternative interpretations. For example, people might simply believe that abilities (or functional traits) are more likely to be innate than disabilities are.

To rule out this kind of alternative interpretation, Knobe and Samuels conducted a second study, in which participants were given no information about the nature of the trait itself. Instead 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.

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").

As predicted by Knobe and Samuels, 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 ( $M = 2.7$ ,  $SD = 1.9$ ). This result lends further support to the view that people's value judgments are impacting their intuitions about innateness.

### 35.5 Innateness Judgments among Scientists

So far we have considered the issue of how naïve subjects deploy the concept of innateness. An obvious next question to ask is whether scientific disciplines that invoke a concept of innateness deploy a distinctively scientific concept or merely redeploy the folk one.

### 35.5.1 Hypotheses

Notice that the pattern of intuitions outlined in Section 35.4 – that moral considerations influence folk innateness judgments but that learning-theoretic considerations do not – is quite antithetical to what one should expect to find in systematic scientific inquiry. The lack of influence exerted by learning-theoretic considerations is surprising because, as already noted, it is no more than a banal truism in the relevant sciences that learned traits are not innate. Indeed, to our knowledge, no hypothesis in psychology – or any related science, for that matter – has *ever* posited traits that are both innate and learned.

The influence exerted by values on folk innateness judgments is similarly antithetical to what one would expect to find in systematic scientific inquiry. There are, of course, many difficult questions about the role of value judgments in scientific inquiry, and different scientists may adopt different views about these questions (Douglas 2009). Nevertheless, it is doubtful that scientists would accept the kind of pattern we find in folk intuitions about innateness. Specifically, it seems remarkable that two scientists who agreed on all the purely descriptive facts in a given case, and disagreed only about their moral significance, should end up disagreeing about any purely scientific issue. Certainly, there are no examples of innateness hypotheses from recent science that exhibit this pattern.

Assuming that this is so, what cognitive processes might explain these departures in a scientific context from the folk innateness judgments elicited in the studies outlined earlier? Three main hypotheses come to mind:

*Overwriting:* Scientific training leads to the elimination of the folk concept of innateness and its replacement by a scientific concept. Among other things, this new concept would emphasize the distinction between learned and non-learned capacities, and it would have no place for the sorts of value considerations outlined earlier.

*Conceptual Addition:* More scientific patterns of innateness judgments depend in part on the acquisition of a new scientific concept, but without the loss of the old one. On this view, scientists hold onto the folk concept but also acquire a scientific one that they use under appropriate conditions – such as those that obtain when doing science.

Both of these proposals assume some augmentation of the scientist’s conceptual resources – presumably as a result of appropriate professional training. But there is a third option, which Knobe and Samuels called the *filtering hypothesis*:

*Filtering:* Scientists never acquire a distinctively scientific concept of innateness. Instead they continue to use the folk concept. However, in arriving at judgments about individual cases scientists do not merely have the innateness concept; 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. So, for example, on the filtering hypothesis, scientists never acquire a new 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, 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.

### 35.5.2 Predictions

Notice that the above three hypotheses make quite different predictions about how the patterns of judgments exhibited by folk and scientists in the relevant fields. If the *overwriting hypothesis* is true, we should expect scientists in the relevant fields only to exhibit a pattern of innateness

judgments in which learning-theoretic considerations exert an influence and moral ones do not.

If the *conceptual addition hypothesis* is correct, we should expect a systematic difference between folk innateness judgments and those made by scientists in the relevant disciplines. Scientists might sometimes use the folk innateness concept, and so exhibit the folk pattern of judgments. But when they have time to reflect carefully on a given case and use their scientific approach (while avoiding the influence of their folk concept), we should expect them to make innateness judgments that are sensitive to learning-theoretic considerations, but not moral ones. In contrast, the folk should *never* exhibit this pattern of sensitivity and insensitivity since, by assumption, they lack the distinct and purely scientific concept of innateness.

Finally, if the *filtering hypothesis* is correct, we should expect scientists' patterns of innateness judgments to sometimes be the same as the folk pattern, and sometimes to exhibit the characteristic scientific pattern. In this regard, it makes the same predictions as the conceptual addition hypothesis. Where the two hypotheses diverge, however, is in their predictions about *folk* judgments. In contrast to the conceptual addition hypothesis, which predicts that folk will not manifest the scientific pattern, the filtering hypothesis allows that non-specialists may, under appropriate circumstances, exhibit the characteristic pattern of judgments. This is because the filtering hypothesis does not presuppose that the scientific pattern in innateness judgments results from the acquisition of special scientific concepts or knowledge. Instead it is wholly compatible with the idea that the relevant filtering processes are available to scientists and non-scientists alike. In view of this, if it were to turn out that people with no relevant scientific training also exhibit the distinctively scientific pattern of judgment, we would have reason to prefer filtering to conceptual addition.

### 35.5.3 Experiments

In order to adjudicate between these competing hypotheses, Knobe and Samuels conducted a large online questionnaire study. Participants were recruited from two very different populations. The first population was a sample of researchers actively working in fields that used the concept of innateness – for example, psychology, linguistics, and biology. The second population – the non-researchers – was composed of people who might be generally scientifically literate but who had no special training in disciplines which deploy 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. In particular, we presented the kinds of vignettes outlined in Section 35.3, but used the “joint-separate” technique (Hsee 1996). Participants were assigned to either a *case-based condition* in which they received just one condition from each of the pairs of cases in a between-subject design, or they were assigned to a *principled condition* in which they received 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; and this, we supposed, might allow us to discriminate between the filtering and conceptual addition hypotheses by facilitating whatever tendency there might be for participants to filter out their intuitive responses.

*Procedure.* Each participant received, in random order, the three questions outlined earlier: the Mother's Milk question, the Trait X question, and the Learning question. 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. 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 35.1.

*Mother’s Milk:* In the case-based version, people gave higher ratings in the abilities condition ( $M = 4.56$ ) than in the disabilities condition ( $M = 3.90$ ). 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. In short: in the principled condition the effect of value – enhanced ability as opposed to disability – was substantially reduced. But equally importantly for our purposes, the populations of researchers and non-researchers behaved in much the same manner. In the case-based condition both groups were *equally* prone to the influence of value; and in the principled condition, the influence of value was similarly reduced.

*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 ( $M = 3.48$ ). In the principled version, this difference was small ( $3.92$ – $3.79$ ), but still significant. In short: in the principled condition the effect of value was substantially reduced. But equally importantly, the populations of researchers and non-researchers behaved in much the same manner. As with the Mother’s Milk question, in the case-based condition both groups were *equally* prone to the influence of value; and in the principled condition, the influence of value was similarly reduced.

*Learning:* In the case-based version, the difference between ratings for the learning condition ( $M = 4.61$ ) and the neuroscience condition ( $M = 4.75$ ) was quite small, but still significant. In the principled condition, there was a more substantial difference between the learning condition ( $M = 4.2$ ) and the neuroscience condition ( $M = 5.3$ ). Moreover, and equally importantly, both the researcher and non-researcher populations exhibited much the same pattern. That is, in the case-based condition neither group’s judgments were much influenced by learning-theoretic considerations, whereas in the principled condition learning-theoretic considerations exerted a far more substantial influence.

*Implications.* What implications do these results have for the relative plausibility of the three hypotheses outlined in Section 35.5.1? First, the data count against the overwrite hypothesis. On this hypothesis, specialist researchers should no longer have any vestige of the folk concept and should, therefore, show the distinctively scientific pattern of judgments, even in the case-based condition. Yet this is not what happens. Rather it is only in the principled condition that researchers’ innateness judgments are influenced learning-theoretic considerations and not moral ones.

**Table 35.1** Descriptive Statistics for Experiment Comparing Scientist and Non-Scientist Innateness Judgments

		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

Second, the data suggest that the conceptual addition hypothesis is implausible. On this hypothesis specialized scientific training is required for the acquisition of new conceptual resources that non-researcher simply do not have. In which case, we should expect researchers to differ from non-researchers in the principled condition. But this is not what the data suggest. Rather, researchers and non-researchers behave alike in both conditions.

This leaves the filtering hypothesis as the one best supported. Indeed, the results are precisely what one would expect if the filtering hypothesis were true. Within the case-based condition, both ordinary folk 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 value 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. Overall, then, the present results suggest that scientists have not replaced or supplemented the folk concept innateness 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.

## 35.6 Conclusion

In this chapter I first set out the evidence for the three-feature account of folk innateness judgments. Second, I argued that this evidence fails to bolster the standard criticism of the innateness concept. Finally, I presented evidence which supports a filtering hypothesis about scientific innateness judgments, and thereby supports one presupposition of the standard criticism – namely that scientists use the same folk innateness concept that non-scientists use. This may appear to be bad news for the role of innateness concept in the sciences. After all, if the folk concept is problematic, then so too is the one that figures in scientific innateness judgments. Even so, the filtering hypothesis also points in the direction of an oft-neglected possibility. Problematic concepts can be used for beneficial theoretical and explanatory purposes in the sciences. This is plausibly true of such concepts as *gene*, *species*, and *element*; and as Fiona Cowie recently argued, the same may be true of the innateness concept as well (Cowie 2009). The filtering hypothesis

suggests one kind of process which may facilitate the beneficial use of problematic concepts in general, and the folk innateness concept, in particular. Under the appropriate circumstances – when we have time to reflect, and the informational context is structured in a manner that facilitates attention to relevant factors and salient general principles – the deleterious inferential effects of a concept can be filtered out and substituted by preferable inferential patterns.

## References

- Alicke, M. 2000. Culpable Control and the Psychology of Blame. *Psychological Bulletin* 126: 556–574.
- Atran, S. 1990. *Cognitive Foundations of Natural History: Towards an Anthropology of Science*. Cambridge and New York: Cambridge University Press.
- Atran, S. 1995. Causal Constraints on Categories. In *Causal Cognition: A Multi-Disciplinary Debate*, edited by D. Sperber, D. Premack, and A. J. Premack, 205–233. Oxford: Clarendon Press.
- Atran, S. 1998. Folk Biology and the Anthropology of Science: Cognitive Universals and Cultural Particulars. *Behavioral and Brain Sciences* 21(4): 547–569.
- Atran, S., D. Medin, E. Lynch, V. Vaparyansky, U.E. Edilberto, and P. Sousa. 2001. Folkbiology Doesn't Come from Folkpsychology: Evidence from Yukatek Maya in Cross-Cultural Perspective. *Journal of Cognition and Culture* 1: 3–42.
- Bateson, P. 1991. Are There Principles of Behavioural Development? In *The Development and Integration of Behaviour: Essays in Honour of Robert Hinde*, edited by P. Bateson, 19–40. Cambridge: Cambridge University Press.
- Beebe, J. R. This volume. Evaluative Effects on Knowledge Attributions.
- Beebe, J. R., and W. Buckwalter. 2010. The Epistemic Side-Effect Effect. *Mind & Language* 25: 474–498.
- Carey, S. 1992. The Origin and Evolution of Everyday Concepts. In *Cognitive Models of Science*, edited by R. Giere and H. Feigl, 15–89. Minneapolis: University of Minnesota Press.
- Carey, S. 2010. *The Origin of Concepts*. Oxford: Oxford University Press.
- Chan, H.-Y., M. Deutsch, and S. Nichols. This volume. Free Will and Experimental Philosophy.
- Cova, F. This volume. The Folk Concept of Intentional Action: Empirical Approaches.
- Cowie, F. 1999. *What's Within? Nativism Reconsidered*. Oxford: Oxford University Press.
- Cowie, F. 2009. Why Isn't Stich an ElimiNativist? In *Stich and His Critics*, edited by Dominic Murphy and Michael A. Bishop, 14–74. Oxford: Wiley-Blackwell.
- Cushman, F., J. Knobe, and W. Sinnott-Armstrong. 2008. Moral Appraisals Affect Doing/Allowing Judgments. *Cognition* 108: 353–380.
- Ditto, P. H., D.A. Pizarro, and D. Tannenbaum. 2009. Motivated Moral Reasoning. In *Psychology of Learning and Motivation: Moral Judgment and Decision Making*, Vol. 50, edited by B. H. Ross (Series Ed.) and D. M. Bartels, C. W. Bauman, L. J. Skitka, and D. L. Medin, 307–338. San Diego, CA: Academic Press.
- Douglas, H. 2009. *Science, Policy, and the Value-Free Ideal*. Pittsburgh, PA: University of Pittsburgh Press.
- Fung, Y. L. 1983. *A History of Chinese Philosophy* (2 vols.). Princeton: Princeton University Press.
- Gelman, S. 2003. *The Essential Child: Origins of Essentialism in Everyday Thought*. New York: Oxford University Press.
- Griffiths, P.E. 2002. What Is Innateness? *The Monist* 85: 70–85.
- Griffiths, P.E., E. Machery, and S. Linquist. 2009. The Vernacular Concept of Innateness. *Mind & Language* 24: 605–630.
- Hitchcock, C., and J. Knobe. 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.
- Knobe, J. 2003. Intentional Action and Side Effects in Ordinary Language. *Analysis* 63: 190–193.
- Knobe, J. 2010. Person as Scientist, Person as Moralizer. *Behavioral and Brain Sciences* 33: 315–329.
- Knobe, J., and R. Samuels. 2013. Thinking Like a Scientist: Innateness as a Case Study. *Cognition* 126 (1): 72–86.
- Linquist, S., E. Machery, P.E. Griffiths, and K. Stotz. 2011. Exploring the Folkbiological Conception of Human Nature. *Philosophical Transactions of the Royal Society B* 366: 444–454.

- Livengood, J., and D. Rose. This volume. Experimental Philosophy and Causal Attributions.
- Machery, E. 2006. The Folk Concept of Intentional Action: Philosophical and Experimental Issues. *Mind & Language* 23: 165–189.
- Machery, E. 2014. Doing Away with Innateness. In *Invited Symposium: Reassessing Innateness*, chaired by R. Samuels, conducted by Central Division APA, Chicago, IL.
- Mameli, M., and P. Bateson. 2011. An Evaluation of the Concept of Innateness. *Philosophical Transactions of the Royal Society B* 366: 436–443.
- Medin, D., and A. Ortony. 1989. Psychological Essentialism. In *Similarity and Analogical Reasoning*, edited by S. Vosniadou and A. Ortony, 179–196. Cambridge: Cambridge University Press.
- Nichols, S., and J. Ulatowski. 2007. Intuitions and Individual Differences: The Knobe Effect Revisited. *Mind & Language* 22: 346–365.
- Phillips, J., and J. Knobe. 2009. Moral Judgments and Intuitions about Freedom. *Psychological Inquiry* 20: 30–36.
- Prior, Arthur. 1967. The Runabout Inference Ticket. In *Philosophical Logic*, edited by Peter Strawson, 38–39. Oxford: Oxford University Press.
- Samuels, R. 2004. Innateness and Cognitive Science. *Trends in Cognitive Sciences* 8: 136–141.
- Sloutsky, V.M., H. Kloos, and A.V. Fisher. 2007. When Looks Are Everything: Appearance Similarity versus Kind Information in Early Induction. *Psychological Science* 18: 179–185.
- Stich, S.P. 1975. *The Idea of Innateness. Inmate ideas*. Los Angeles: University of California Press.
- 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., and J. Phillips. 2011. The Paradox of Moral Focus. *Cognition* 119: 166–178.