Journal of Educational Psychology
2013, Vol. 105, No. 1, 199–212


0022-0663/13/$12.00 DOI: 10.1037/a0030282

# Epistemic Similarities Between Students' Scientific and Supernatural Beliefs

Andrew Shtulman
Occidental College

The evidential support for scientific claims is quantitatively and qualitatively superior to that for supernatural claims, yet students may not appreciate this difference in light of the fact that both types of claims are learned in similar ways (through testimony rather than firsthand observation) and perform similar functions (explaining observed phenomena in terms of unobservable entities). The present study addressed this issue by comparing students' scientific beliefs with their supernatural beliefs along 4 dimensions of epistemic import: personal confidence, perceived consensus, means of justification, and openness to revision. Participants' scientific beliefs were strongly differentiated from their supernatural beliefs along the dimensions of confidence and consensus but only weakly differentiated along the dimensions of justification and revision, particularly for participants with (a) higher levels of supernatural belief and (b) lower levels of understanding of the nature of science. Moreover, participants' confidence in both types of beliefs was associated with their consensus estimates but not with their ability to cite evidence in support of, or potentially in conflict with, those beliefs. These findings imply that many students' scientific beliefs are qualitatively similar to their supernatural beliefs, despite self-perceptions to the contrary.

*Keywords:* epistemology, belief, religion, science understanding, science education

> The analogy between the magical and the scientific conceptions of the world is close. In both of them the succession of events is assumed to be perfectly regular and certain, being determined by immutable laws, the operation of which can be foreseen and calculated precisely; the elements of caprice, of chance, and of accident are banished from the course of nature. Both of them open up a seemingly boundless vista of possibilities to him who knows the causes of things and can touch the secret springs that set in motion the vast and intricate mechanism of the world. (Frazer, 1922/1998, p. 45)

In the *Golden Bough*, anthropologist Jonathan Frazer (1922/1998) famously argued that scientific thinking and magical thinking stem from the same motivation: the need to explain, predict, and control the natural world. Frazer saw this motivation as manifesting itself in different procedures but ultimately yielding similar outcomes. That is, the products of science and magic—causal models of natural phenomena and procedures for manipulating those phenomena—share many commonalities even if the practices of science and magic do not. Both provide frameworks for interpreting everyday observation and experience; both posit unobservable entities as the causes of observable phenomena; and both extend, or even defy, early developing intuitions about the kinds of entities that exist and the kinds of interactions those entities engage in (see McCauley, 2000). For instance, just as scientists explain chemical reactions in terms of subatomic particles that (a) cannot be seen, (b) cannot be directly interacted with, and (c) possess properties that are inconsistent with the properties of matter in general, theists explain the orderly nature of the universe in terms of a divine agent that (a) cannot be seen, (b) cannot be directly interacted with, and (c) possesses properties that are inconsistent with properties of agents in general.

Similarities in the *content* of scientific and supernatural claims stand in marked contrast to differences in the *process* by which those claims are generated and evaluated. As many prominent scientific bodies have noted (e.g., American Association for the Advancement of Science, 2002; Inter-Academy Panel, 2006; National Academy of Sciences, 1999), only scientific explanations generate testable hypotheses; only scientific explanations are supported by experiments with observable outcomes; and only scientific explanations are imminently revisable and potentially defeasible. Professional scientists see such distinctions as clear and obvious. The general public, however, is more ambivalent, as evident from recent debates over whether intelligent design should be taught in the public schools as a scientific alternative to evolution by natural selection (Brockman, 2006; Lombrozo, Shtulman, & Weisberg, 2006) and whether climate change is actually occurring (Leiserowitz, 2005, 2006). Differences in the epistemic foundations of scientific and supernatural claims, although highly salient to those who generate such claims (i.e., practitioners of science), may not be so salient to those who learn them in passing (i.e., students of science).

To what extent do students appreciate the vast difference in evidential support for scientific claims versus supernatural claims? Although no studies have addressed this question directly, research on students' understanding of science as a *method* or *process* suggests that students' epistemologies of science are generally quite shallow (Carey, Evans, Honda, Jay, & Unger, 1989; Chinn &

This article was published Online First October 1, 2012.
This research was supported by National Science Foundation Grant DRL-0953384 awarded to Andrew Shtulman.
Correspondence concerning this article should be addressed to Andrew Shtulman, Department of Psychology, Occidental College, Los Angeles, CA 90041. E-mail: shtulman@oxy.edu

Brewer, 2001; Klahr, Fay, & Dunbar, 1993; Smith, Maclin, Houghton, & Hennessey, 2000; Smith & Wenk, 2006). For instance, research by Kuhn and colleagues (Kuhn, 1991; Kuhn, Amsel, & O'Loughlin, 1988) has shown that students tend to conflate theory and data, treating data as an elaboration of a theory rather than something that exists independent of it and bears on its validity. Research by Lederman and colleagues (Lederman, Abd-El-Khalick, Bell, & Schwartz, 2002; Lederman & O'Malley, 1990) has shown that students consistently misinterpret empirical support for definitive proof, viewing the purpose of an experiment as "proving a hypothesis true" rather than testing, and thus potentially falsifying, the hypothesis. And research by Schauble and colleagues (Schauble, Glaser, Duschl, Schulze, & John, 1995; Schauble, Klopfer, & Raghavan, 1991) has shown that students tend to hold an "engineering view" of experimentation, construing experiments as chiefly about solving problems and making discoveries rather than about testing hypotheses and drawing inferences. Even individuals who hold doctorates in the humanities tend to believe (a) that scientists abide by a single, deterministic method of inquiry, (b) that scientists develop their ideas solely through observation and not also through inference, and (c) that scientists who study the same data will inevitably arrive at the same conclusions (Lederman et al., 2002).

The present study begins to explore students' understanding of science as a body of knowledge, generated by, and responsive to, empirical data, as established from the perspective of how students justify their scientific beliefs. Although there already exists much research on students' understanding of domain-specific concepts (Carey, 2009; Slotta & Chi, 2006; Vosniadou, 1994), that research has not broached the question of what students know about the *evidential foundations* of such concepts. Much is known, for instance, about the cognitive constraints that shape students' understanding of evolutionary phenomena, like adaptation and speciation (Novick, Shade, & Catley, 2010; Shtulman, 2006; Shtulman & Schulz, 2008), but much less is known about the epistemic considerations behind students' acceptance (or denial) of the actual occurrence of evolution. Likewise, much is known about the cognitive constraints that shape students' understanding of material phenomena, like solvency and buoyancy (Au, 1994; Nakhleh & Samarapungavan, 1999; Smith, 2007), but much less is known about the epistemic considerations behind students' acceptance (or denial) of the actual existence of subatomic particles.

To address this gap in the literature, I turned to Hofer and Pintrich's (1997) analysis of the key determinants of epistemological understanding, or students' conceptions of the nature of knowledge and the nature of knowing. Hofer and Pintrich reviewed six different models of epistemological understanding—ranging from Baxter Magolda's (1992) "Epistemological Reflection" model to King and Kitchener's (1994) "Reflective Judgment" model to Kuhn's (1991) "Argumentative Reasoning" model—and identified two dimensions that underlie the majority of them: *certainty of knowledge* and *justification for knowing*. Certainty of knowledge refers to conceptions of the stability and accuracy of one's knowledge, ranging from fixed and absolute, acquired via direct access to universal truths, to tentative and changing, open to new arguments and new interpretations. Justification for knowing refers to the process by which knowledge claims are evaluated, ranging from an unquestioning deference to authority to the critical examination of evidence. I used this analysis in the present study as a point of departure from previous work on epistemological understanding by defining tasks that measured each construct in the context of particular, domain-specific beliefs. Certainty of knowledge was measured by asking participants for an explicit rating of their confidence in the existence of a particular entity as well as by asking participants to reflect on what evidence, if any, might persuade them to change their mind about the existence of that entity. Justification for knowing was measured by asking participants for an explicit justification of their belief as well as by comparing the nature of those justifications with other independent measures of participants' epistemic commitments (described below).

Participants' views regarding the certainty and justifiability of their scientific beliefs were assessed not in isolation but in relation to their views regarding claims of a less evidential nature: supernatural claims. Supernatural claims constituted an ideal basis of comparison for both conceptual and methodological reasons. Conceptually, they provided a close approximation to scientific claims in their form (ideas accepted primarily on the basis of testimony rather than firsthand observation; see Harris & Koenig, 2006) as well as their function (explaining and predicting everyday phenomena; see Legare, Evans, Rosengren, & Harris, 2012). Methodologically, they provided a way of comparing students' reasons for endorsing well-substantiated concepts, like electrons and X-rays, to their reasons for endorsing not-so-well-substantiated concepts, like angels and ghosts, within the very same student, thus increasing the sensitivity of that comparison. Four questions were of particular interest.

First, are students' reasons for endorsing scientific claims substantively different from their reasons for endorsing supernatural claims? Previous research on belief justification has focused either on individual differences in how people justify a single belief, like belief in evolution (Brem, Ranney, & Schindel, 2003) or belief in God (Gottlieb, 2007), or on individual differences in how people justify their positions on nonscientific issues, like the causes of poverty (Kuhn, 1991) or the acceptability of the death penalty (Kuhn & Udell, 2003). Findings from the second line of research indicate that students do not typically privilege evidence over other types of considerations when justifying a position. Instead, they support their positions with reiterations or elaborations of the position itself. For instance, a student charged with the task of justifying a particular position on why children fail at school ("parental involvement") is more likely to reiterate the position ("parents need to be involved in their child's education") than to identify a relevant source of empirical data ("students who have parents who care about their education probably have higher GPAs than students who don't"). This inattentiveness to evidence in the context of argumentation suggests that students may be inattentive to evidence in the context of belief endorsement as well.

Second, are students' reasons for revising their scientific beliefs substantively different from their reasons for revising their supernatural beliefs? Within the science education literature, the topic of belief revision has been studied primarily from the perspective of how students analyze and interpret belief-inconsistent data. Chinn and Brewer (1998), for instance, have identified a number of common strategies for dealing with belief-inconsistent data, including ignoring the data, rejecting the data, holding the data in abeyance, reinterpreting the data, and (in rare occasions) accepting the data as a cause for changing one's beliefs. Which strategy a student uses depends primarily on how well the student can explain the anomalous data in terms of a preexisting causal model (Chinn & Brewer, 2001; Koehler, 1993; Koslowski, Marasia, Chelenza, & Dublin, 2008), though strategies that preserve one's causal models are both more numerous and more frequent than those that do not. The implication of these findings for the research at hand is that, although students should be more sensitive to the possibility of anomalous data when evaluating scientific beliefs than when evaluating supernatural beliefs (which are not, presumably, based on data), they may not demonstrate such sensitivity if their scientific beliefs are based on something other than empirical considerations.

Third, is students' confidence in the validity of scientific claims grounded in different considerations than their confidence in the validity of supernatural claims? In pilot studies, students proved extremely confident in their scientific beliefs, typically claiming to be 100% certain of the existence of everything from electrons to black holes, yet the source of that confidence was unclear. Although, on one hand, it may stem from knowledge of the evidence supporting the entity's existence, it may, on the other hand, stem from more superficial considerations, like the length of time one has believed in the existence of that entity or perceptions of how many *other people* believe in the existence of that entity. Consistent with the latter possibility, Harris, Pasquini, Duke, Asscher, and Pons (2006) found that young children's confidence in the existence of unobservable entities tracked their perceptions of consensus surrounding those entities. In other words, children were highly confident of the existence of entities they claimed everyone believes to exist, like germs and oxygen, but only moderately confident of the existence of entities which they claimed only some people believe to exist, like Santa Claus and God. This correspondence between children's beliefs and their perception of others' beliefs indicates that knowledge of the evidential basis of scientific claims is not a prerequisite for believing in those claims, though such knowledge may ultimately supplant children's initial reasons for belief.

Fourth, does variation in the justification and evaluation of specific scientific beliefs track variation in understanding of the nature of science (NOS)? Research on NOS understanding has shown that students misunderstand the *process* of science in ways that may be relevant to their understanding of the *products* of science, as noted above. Moreover, NOS understanding has been shown to predict many disparate, yet scientifically important, skills, like the ability to integrate formal scientific principles into one's everyday understanding of domain-specific phenomena (Songer & Linn, 1991), the ability to identify reasons for scientific controversy and means for potentially resolving them (Smith & Wenk, 2006), and the acceptance of counterintuitive scientific ideas, like evolution and common descent (Lombrozo, Thanukos, & Weisberg, 2008). That said, science is a multifaceted topic, and there are many facets of NOS understanding tapped by standard, survey-based assessments that seem to run orthogonal to the kinds of conceptions probed here (e.g., the inferential gap between data and theory, the nonlinear trajectory of scientific investigations). It is thus an open question whether an understanding of how science operates as a discipline is related to an understanding of the evidential foundations of particular scientific claims.

Taken together, these four questions subserved the broader objective of exploring students' epistemologies of science by characterizing the epistemic commitments implicit in students' scientific beliefs, as opposed to the epistemic commitments they can explicitly articulate when reflecting on the nature and origin of scientific knowledge. This change in focus proved fruitful methodologically in that it provided a substantively different, yet conceptually complementary, view of students' epistemologies of science from those documented previously. It also proved fruitful theoretically in that it highlighted the extent to which students' scientific beliefs are qualitatively similar to their supernatural beliefs, despite normative prescriptions to the contrary (e.g., Brockman, 2006; Dawkins, 2009).

Although participants' supernatural beliefs were assessed mainly as a metric of comparison for their scientific beliefs, participants' treatment of supernatural claims is potentially of interest in its own right, given potential pedagogical implications of overlapping scientific and supernatural worldviews (Blancke, De Smedt, De Cruz, Boudry, & Braeckman, 2011; Legare et al., 2012; Lindeman, Svedholm, Takada, Lonnqvist, & Verkasalo, 2011). On one hand, educators might find it useful to embrace the overlap, encouraging students to evaluate competing scientific and supernatural explanations of the same phenomena (e.g., evolution vs. God, viruses vs. witchcraft) in terms of evidence and inquiry and thereby come to recognize the degree to which scientific explanations are more empirically valuable than supernatural ones. On the other hand, they may find it more useful to disregard or dispute the overlap, deeming it unproductive to challenge beliefs based primarily on nonempirical considerations or to teach "controversies" that have been settled in the scientific community for decades, if not centuries. Both approaches are defensible, but both approaches might not be equally effective. The present study helps to shed light on this issue by detailing, for the first time, the *epistemic* overlap between students' scientific and supernatural beliefs. The findings, reviewed below, suggest that the approach of explicitly contrasting the two types of beliefs might resonate best with students' underlying epistemologies.

## Method

### Participants

One hundred forty college undergraduates participated in the study for course credit in an introductory psychology class. Approximately half were recruited from a large, urban university in the Northeastern United States (Harvard University) and half from a small, urban college in the Southwestern United States (Occidental College). Preliminary analyses revealed no significant differences between the two groups on any of the measures reported below, so they were pooled together.

### Procedure

Participants completed a questionnaire that probed their beliefs about six scientific entities (black holes, electrons, evolution, fluoride, genes, X-rays) and 12 supernatural entities (angels, fate, ghosts, God, Heaven, Hell, karma, precognition, reincarnation, Satan, souls, and telepathy). Pilot data confirmed that, when presented with this particular selection of entities, undergraduates tended to endorse an equal number of entities from the two domains, that is, typically all six scientific entities and around six of the 12 supernatural entities. For each entity, participants were asked five questions: (a) whether they believed the entity exists; (b) how confident they were of that belief; (c) how many other Americans hold the same belief; (d) why they believe the entity exists; and (e) what evidence, if any, might persuade them to change their belief. Responses to these questions are henceforth referred to as *existence judgments*, *confidence ratings*, *consensus estimates*, *belief justifications*, and *belief refutations*, respectively.

Participants' belief justifications and belief refutations were analyzed using coding schemes described below. These schemes were applied to the entire data set by two independent coders: the author, who created them, and a research assistant, who was instructed on how to apply them but was not involved in their creation. Intercoder agreement was 95% for the belief justifications (Cohen's $\kappa$ = .94) and 94% for the belief refutations (Cohen's $\kappa$ = .92), and all disagreements were resolved through discussion.

Following the questionnaire, a subset of participants ($n$ = 65) were also administered a seven-item, multiple-choice assessment of their NOS understanding. This assessment, which can be found in the Appendix, was modeled on those used by Carey et al. (1989) and Lederman et al. (2002) and covered (a) the nature of an experiment, (b) the nature of a theory, (c) the role of empirical data, and (d) the role of inference. Given the assessment's length and format, it was not intended to constitute a comprehensive measure of all aspects of NOS understanding. Rather, it was intended to provide a preliminary look at whether, and how, NOS understanding relates to the justification and evaluation of specific scientific beliefs.

Table 1
*Mean Existence Judgments (Range = 0–1), Confidence Ratings (Range = 1–7), and Consensus Estimates (Range = 1–7) for the Six Scientific Items and 12 Supernatural Items*

| Item | Existence judgments | Confidence ratings | Consensus estimates |
|---|---|---|---|
| Electrons | .99 | 6.5 | 6.0 |
| Fluoride | .99 | 6.8 | 6.3 |
| Genes | .99 | 6.8 | 6.4 |
| X-rays | .98 | 6.9 | 6.6 |
| Evolution | .96 | 6.3 | 4.7 |
| Black holes | .95 | 5.8 | 5.1 |
| Souls | .81 | 5.9 | 5.4 |
| God | .70 | 5.8 | 5.2 |
| Karma | .68 | 5.5 | 4.0 |
| Heaven | .59 | 5.6 | 5.2 |
| Fate | .53 | 5.4 | 4.5 |
| Angels | .51 | 5.4 | 3.9 |
| Ghosts | .44 | 4.6 | 3.6 |
| Hell | .41 | 5.1 | 4.7 |
| Precognition | .41 | 4.9 | 3.1 |
| Telepathy | .36 | 4.7 | 3.0 |
| Satan | .32 | 5.8 | 4.5 |
| Reincarnation | .32 | 4.8 | 3.0 |

## Results

### Existence Judgments

The first question participants answered about each entity was whether or not they believed in its existence. "Yes" responses were assigned 1 point, and "No" responses were assigned 0 points. Participants' mean existence judgments are displayed in Table 1. On average, participants endorsed the existence of 5.9 scientific items (or 98%) and 6.1 supernatural items (or 51%). Rates of endorsement for the supernatural items were generally lower than those obtained in national surveys of supernatural belief but were still ordinally similar in that religious items, like God and angels, were endorsed more frequently than paranormal items, like ghosts and telepathy (Moore 2005; Winseman, 2004). It should be noted that only items judged existent were used to compute mean scores for a given participant, as only those items were relevant to the question of how the *endorsement* of scientific claims compares with the *endorsement* of supernatural claims. This procedure meant that participants who denied the existence of all 12 supernatural entities ($n$ = 9) had to be excluded from analyses that compared scientific items with supernatural items, but those participants were retained for all other analyses.

### Confidence Ratings

Participants rated their confidence in the existence of each entity on a scale ranging from 1 (*not confident*) to 7 (*100% confident*). Participants tended to be extremely confident in their endorsements, selecting "7" significantly more often than any other rating (46% of all selections, binomial $p < .001$). They also tended to exhibit greater confidence in the scientific entities than in the supernatural entities, as shown in Table 1. Not only was the mean rating for all scientific items significantly greater than the mean rating for all supernatural items ($M$ = 6.4 vs. $M$ = 5.2), $t(130)$ = 11.85, $p < .001$, but the ratings for individual scientific items were almost always greater than the ratings for individual supernatural items. These ratings closely corresponded to the existence judgments, with participants exhibiting greater confidence for entities judged existent by a larger portion of the sample: correlation across items, $r(18) = .88$, $p < .001$.

### Consensus Estimates

After selecting a confidence rating, participants estimated the number of Americans who would agree with their existence judgment on a scale ranging from 1 (*1 out of 7*) to 7 (*7 out of 7*). Mean consensus estimates for each entity are displayed in Table 1. Similar to the confidence ratings, consensus estimates for the scientific items were significantly greater than those for the supernatural items ($M$ = 5.8 vs. $M$ = 4.4), $t(130) = 16.08$, $p < .001$. At the item level, mean consensus estimates were highly correlated with mean existence judgments, $r(18) = .83$, $p < .001$, indicating that participants' estimates were at least partially veridical.

### Belief Justifications

Participants provided a total of 1,682 justifications for their existence judgments (824 for scientific items and 858 for supernatural items), and these justifications were sorted into one of four categories in accordance with the decision tree in Figure 1. The first sort was between responses that served as a justification for belief—that is, responses that actually answered the question "Why do you believe in the existence of [entity]?"—and responses

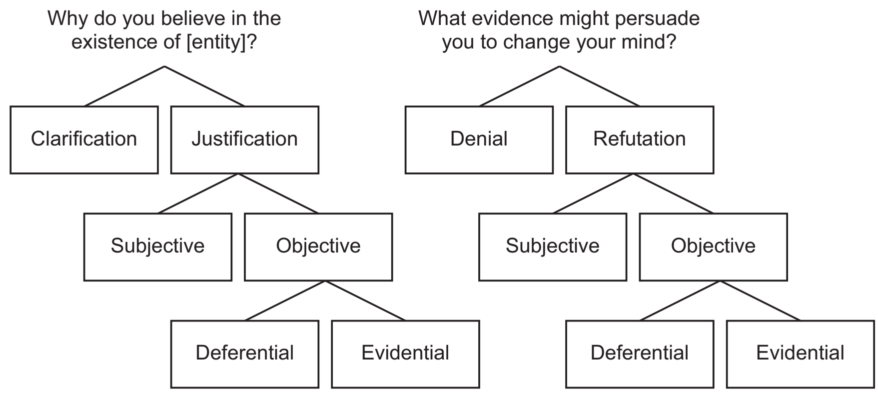


*Figure 1.* Decision trees for coding belief justifications and belief refutations.

that merely clarified or qualified the nature of the belief (e.g., "I believe that organisms adapt to their environment, but not that we all come from one common being"; "I believe in the presence of those who have passed away, and I suppose this is what you would call an angel"). The latter are treated simply as nonjustifications.

The second sort was between responses that cited an external, objective reason for belief and responses that cited a more personal, subjective reason for belief. This latter category, termed *subjective justifications*, included appeals to intuition (e.g., electrons must exist because "they make rational sense"; precognition must exist because "it sounds plausible"), appeals to volition (e.g., Heaven must exist because "I like to think that my loved ones are going there"), and appeals to a direct encounter with the entity in question (e.g., genes must exist because "we looked at DNA in bio class once"; angels must exist because "I've seen one"). The rationale for coding appeals to a direct encounter as a subjective justification was twofold. First, these justifications were predicated on a particular experience or point of view not necessarily shared with other individuals and thus not necessarily persuasive to other individuals. Second, these justifications presupposed a degree of interpretation that the objective justifications did not. DNA, for instance, is observable, but the theoretical construct of a "gene" is not. Likewise, coincidental events are observable, but the theoretical construct of "fate" is not. That said, coding these justifications as a form of evidence would not have radically changed the results, as they comprised a relatively small—and relatively equal—portion of justifications in both domains (12% of the scientific justifications, 11% of the supernatural justifications).

The third sort was between two types of objective justifications: those that referenced independently verifiable facts that support (or are believed to support) the existence of the entity in question and those that referenced the source of one's belief without referencing any factual or conceptual considerations relevant to the legitimacy of that source. The first type of objective justification, termed *evidential*, included information about an entity's observable properties (e.g., genes must exist because "they can be sequenced and manipulated"; souls must exist because "there is an inexplicable loss of 22 grams at the time of death"), causal effects (e.g., black holes must exist because "they cause a lack of light in certain regions of space"; God must exist because "the Universe had to come from somewhere") or historical record (e.g., evolution must exist because "there are fossils for past species that have similar characteristics to present day animals"; ghosts must exist because "stories about them are universal"). The second type of objective justification, termed *deferential*, included appeals to authority (e.g., black holes must exist because "I trust my physics teacher Mr. Murray"; Hell must exist because "it's in the Bible"), appeals to instruction (e.g., electrons must exist because "I've learned about them in most of the science classes I've taken"; God must exist because "it's what I was taught"), and appeals to an overall worldview consistent with, or dependent on, the entity in question (e.g., fluoride must exist because "I'm a chemistry major"; angels must exist because "I'm Muslim; angels exist by default").

Overall, the coding procedure could be seen as a way of distilling evidential justifications from other forms of justification, as evidential justifications are clearly privileged in the domain of science (Kuhn, 1991; Toulmin, 1964). Evidential justifications are not, however, the only kind of justification that factor into scientific arguments; deference to the work of others plays an important role as well (Keil, 2010; Kitcher, 2001). The following analyses thus focus on both kinds of justifications—evidential justifications and deferential justifications—examining how they vary, relative to one another, within and across domains. What those patterns indicate about participants' epistemological understanding of scientific claims is an issue considered more fully in the Discussion.

The relative frequencies of each type of justification to total justifications are displayed in Table 2 by domain. Overall, participants provided deferential justifications more than any other type of justification in both domains. Evidential justifications were next most frequent in the scientific domain, whereas subjective justifications were next most frequent in the supernatural domain. Paired samples $t$ tests revealed that two types of justifications were significantly more common in the scientific domain than in the supernatural domain: evidential justifications, $t(130) = 5.49$, $p < .001$; and deferential justifications, $t(130) = 3.85$, $p < .001$. Nevertheless, the size of those differences were small—namely, a 16% difference in the provision of evidential justifications and a 12% difference in the provision of deferential justifications. The justification profiles for the two domains were thus more similar than what might be expected from the sharp difference in confidence ratings and consensus estimates. Indeed, the size of that difference, measured in terms of Cohen's $d$ (Cohen, 1988), was quite large for the confidence ratings ($d = 1.38$) and consensus

Table 2
*Mean Proportion (+SE) of Justification Types and Refutation Types in Each Domain and the Paired Samples Correlations Between Them*

| Measure | Scientific domain | Supernatural domain | Correlation |
|---|---|---|---|
| Evidential justifications | .29 (.02) | .13 (.02) | .28** |
| Deferential justifications | .53 (.02) | .41 (.03) | .20* |
| Subjective justifications | .17 (.01) | .31 (.02) | .09 |
| Nonjustifications (clarifications) | .01 (.00) | .14 (.00) | −.09 |
| Evidential refutations | .30 (.02) | .20 (.03) | .46** |
| Deferential refutations | .39 (.03) | .21 (.02) | .23** |
| Subjective refutations | .10 (.01) | .28 (.03) | .31** |
| Nonrefutations (denials) | .21 (.02) | .30 (.03) | .19* |

* $p < .05$. ** $p < .01$.

estimates ($d = 1.50$) but quite small for the provision of evidential justifications ($d = 0.48$) and deferential justifications ($d = 0.33$).

This between-domain similarity in justification profiles can also be seen at the level of individual participants. The more often a participant provided evidential justifications in one domain, the more often he or she provided evidential justifications in the other domain as well, $r(131) = .28$, $p < .01$. Likewise, the more often a participant provided deferential justifications in one domain, the more often he or she provided deferential justifications in the other domain, $r(131) = .20$, $p < .05$. These correlations are explored more extensively below.

## Belief Refutations

Belief refutations ($n = 1,682$) were coded in a similar manner to the belief justifications, as shown in Figure 1. First, responses in which participants cited considerations that might actually change their mind were differentiated from responses in which participants denied that *anything* could change their mind (e.g., "there is no evidence that could affect my belief in fate"; "Nothing at this point can dissuade me from the idea of evolution"). This latter set of responses was termed *denials* and was treated as distinct from the remaining set of responses.

Second, responses that cited external, objective reasons for changing one's mind were differentiated from responses that cited personal, subjective ones. These "subjective refutations" typically took the form of a hypothetical event that, if experienced, would call the entity's existence into question (e.g., one's belief in X-rays would be challenged "if I found out all the X-rays I underwent were staged"; one's belief in Hell would be challenged "if I died and wasn't punished for my sins"). Responses that indicated a willingness to change one's mind but a vagueness or uncertainty about what might incite that change (e.g., "maybe I'll feel differently one day"; "I could be persuaded otherwise") were also included in this category.

Third, responses that cited substantive facts or ideas that would challenge the belief in question (labeled as *evidential refutations*) were differentiated from those that cited *sources* of evidence but not the evidence itself (labeled as *deferential refutations*). Evidential refutations included the possibility of discovering anomalous data (e.g., one's belief in electrons would be challenged "if an atom was found without them"; one's belief in karma would be challenged "if bad people started experiencing good things") or the possibility of discovering an alternative causal model of the phenomena of interest (e.g., one's belief in fluoride would be challenged "if a new scientific model was heavily endorsed that could explain the building blocks of life without using the elements in the periodic table"; one's belief in souls would be challenged "if science could find a way to explain why there is life at all and how individuality is created in terms of thinking and feeling"). Deferential refutations, however, referenced an informant, or group of informants, whose change of mind was sufficient to incite a personal change of mind (e.g., one's belief in X-rays would be challenged "if a bunch of scientists got together and proved they didn't exist"; one's belief in Satan would be challenged "if the Church said it did not exist"). Also included in this category were responses that referenced a general system of beliefs that, if abandoned, would result in a change to the specific belief at hand (e.g., one's belief in evolution would be challenged by "becoming extremely religious"; one's belief in reincarnation would be challenged by "more exposure to alternative beliefs").

The overall pattern of belief refutations was similar to the overall pattern of belief justifications, as shown at the bottom of Table 2. Deferential refutations were provided more frequently than evidential refutations across domains ($M = 0.34$ vs. $M = 0.25$), $t(139) = 2.17$, $p < .05$, and both types of refutations were provided significantly more often in the scientific domain than in the supernatural domain: evidential refutations, $t(130) = 3.55$, $p < .001$; deferential refutations, $t(130) = 5.54$, $p < .001$. However, the size of the domain difference was once again quite small—that is, an 11% difference in the overall frequency of evidential refutations ($d = 0.33$) and an 18% difference in the overall frequency of deferential refutations ($d = 0.48$)—and the response patterns of individual participants were once again consistent across domains: evidential refutations, $r(131) = .46$, $p < .001$; deferential refutations, $r(131) = .23$, $p < .01$.

Perhaps the most noteworthy finding to emerge from this analysis was the high rate at which participants denied that anything could change their mind about the existence of the entities in question. Participants provided this type of response 30% of the time in the supernatural domain and 21% of the time in the scientific domain, with over half (55%) providing this type of response for at least one scientific entity. Moreover, the provision of this type of response was correlated across domains, $r(131) = .19$, $p < .05$, implying that the more participants tended to view their supernatural beliefs as independent of (or impervious to) evidence, the more they tended to view their scientific beliefs in the same way.

## Interrelations Among Epistemic Dimensions

Thus far, it has been shown that participants (a) exhibited greater confidence in their scientific beliefs, (b) perceived greater consensus surrounding their scientific beliefs, (c) cited (slightly) more evidence in support of their scientific beliefs, and (d) cited (slightly) more counterevidence to their scientific beliefs. These results indicate that participants' understanding of scientific claims is not *identical* to their understanding of supernatural claims. Participants do, in fact, differentiate the two, consistent with the finding that even young children differentiate the two (Harris et al., 2006). The extent of that differentiation, however, is quite limited, as demonstrated by the following analyses of the interrelations among the measures of interest.

**Interrelations between justifications and refutations.** The provision of evidential responses was not widespread. In the scientific domain, only 22 participants (or 16%) provided evidential justifications as their modal justification type, and only 29 participants (or 21%) provided evidential refutations as their modal refutation type. The similarity in these proportions is not coincidental. Participants who provided a high number of evidential justifications also tended to provide a high number of evidential refutations, and they tended to do so in both the scientific domain and the supernatural domain, as indicated by the correlations in Table 3. Indeed, the correlations in Table 3 show that this pattern was true of participants' deferential responding as well; participants who provided a high number of deferential justifications also tended to provide a high number of deferential refutations, regardless of the domain. Participants thus appeared to differ in their approach to the task as a whole, with some focusing consistently on providing evidence in support of, or in conflict with, their beliefs and others focusing consistently on naming the *sources* of such evidence but not the evidence itself. These findings complement those of Sá, Kelley, Ho, and Stanovich (2005), who also document consistent individual differences in students' tendency (or ability) to cite evidence in support of their beliefs, albeit using different methods and different analyses.

**Interrelations between justifications, refutations, and supernatural belief.** As a group, participants provided deferential justifications nearly twice as often as they provided evidential justifications in the scientific domain ($M$ = 0.53 vs. $M$ = 0.29), $t(139) = 5.59$, $p < .001$. As individuals, however, this difference varied greatly, with some participants providing only deferential justifications and others providing only evidential justifications. This variance was compared with two other measures of epistemic import: levels of supernatural belief and levels of NOS understanding (reported below). The results of the first comparison are displayed in Figure 2, which depicts the mean proportion of evidential and deferential responses provided by participants in each quartile of the distribution of existence judgments for the supernatural entities (Quartile 1 = 0–2 supernatural entities judged existent; Quartile 2 = 3–6; Quartile 3 = 7–8; Quartile 4 = 9–12).

Table 3
*Correlations Within and Across Domains Regarding the Provision of Evidential Responses (Top Panel) and Deferential Responses (Bottom Panel)*

| Type | Measure | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| Evidential | 1. Scientific justifications | — | .51** | .28** | .23** |
| | 2. Scientific refutations | | — | .29** | .46** |
| | 3. Supernatural justifications | | | — | .48** |
| | 4. Supernatural refutations | | | | — |
| Deferential | 1. Scientific justifications | — | .40** | .20* | .27** |
| | 2. Scientific refutations | | — | .21* | .23** |
| | 3. Supernatural justifications | | | — | .02 |
| | 4. Supernatural refutations | | | | — |

* $p < .05$. ** $p < .01$.

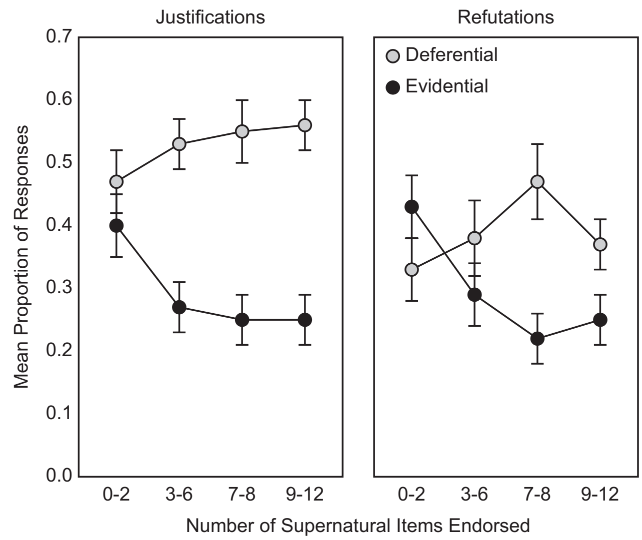


*Figure 2.* The mean proportion (+*SE*) of evidential and deferential refutations for the scientific items provided by participants of varying supernatural beliefs.

Participants in all quartiles appeared to provide the same number of deferential responses, but participants in the first quartile appeared to provide more evidential responses than those in the other three quartiles. Statistical confirmation of this trend was obtained with regression analyses in which the difference between the number of evidential responses provided and the number of deferential responses provided was regressed against the number of supernatural entities judged existent. These analyses were significant for both the justification data ($R = 0.19$), $F(1, 139) = 5.09$, $p < .05$, and the refutation data ($R = 0.20$), $F(1, 139) = 5.53$, $p < .05$, indicating that participants who were particularly skeptical of *supernatural* claims tended to cite more evidence when reasoning about *scientific* claims.

**Interrelations between justifications, refutations, and NOS understanding.** Just as the provision of evidential and deferential responses differed by levels of supernatural belief, they also differed by levels of NOS understanding (for the 65 participants who completed the NOS assessment, at least). Participants' scores on the NOS assessment ranged from 1 to 6, with approximately half providing scores between 1 and 3 and half providing scores between 4 and 6. Figure 3 displays participants' justification and refutation profiles for the scientific items as a function of their NOS score (1–3 vs. 4–6). Participants with high NOS scores tended to provide more evidential responses than those with low NOS scores, and participants with low NOS scores tended to provide more deferential responses than those with high NOS scores. Regression analyses confirmed that the difference between the number of evidential responses provided and the number of deferential responses provided was significantly predicted by NOS scores for both the justification data ($R = 0.33$), $F(1, 64) = 7.78$, $p < .01$, and the refutation data ($R = 0.34$), $F(1, 64) = 8.35$, $p < .01$. Thus, participants' *explicit* understanding of the nature of science was predictive of their *implicit* understanding, as measured

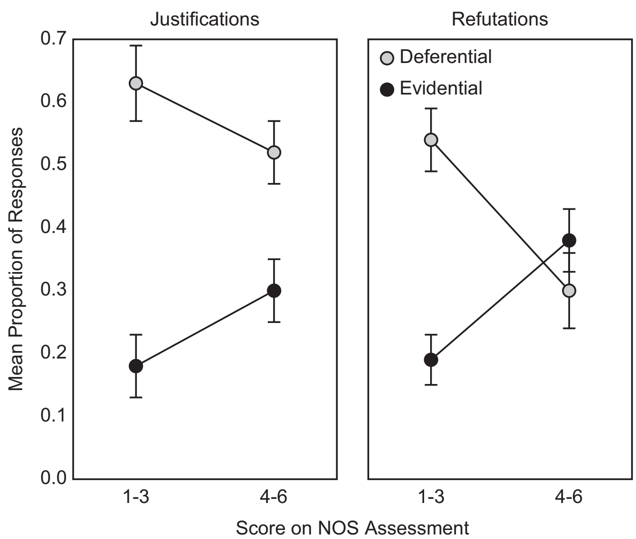


*Figure 3.* The mean proportion (+*SE*) of evidential and deferential justifications for the scientific items as a function of participants' score on the nature of science (NOS) assessment.

by differences in how they justified and evaluated their scientific beliefs.

**Interrelations between justifications, refutations, and confidence.** As noted above, the difference in participants' confidence ratings between the scientific and supernatural domains was substantially larger than the difference in their patterns of justification and refutation, implying that the former may be largely independent of the latter. To test this hypothesis—that beliefs held with high confidence were no more likely to be justified (or refuted) evidentially than beliefs held with lower confidence—participants' confidence ratings were directly compared to their evidential justifications and evidential refutations. This analysis was accomplished by comparing the mean frequency of evidential responses for items assigned a confidence rating of 7 (or "100% confident") with that for items assigned a confidence rating of 6 or lower. Seven was chosen as the cutoff point because (a) it was participants' modal confidence rating, comprising 46% of all ratings, and (b) establishing a cutoff at this point allowed for the greatest number of within-participant comparisons ($n$ = 116 for the scientific items and $n$ = 71 for the supernatural items), because not all participants used the full range of the rating scale. Indeed, 19% of participants provided the same confidence rating for all scientific items, and 26% provided the same confidence rating for all supernatural items.

Despite this problem, a consistent pattern emerged among the participants who did exhibit variation in their confidence ratings: Such ratings were associated neither with the provision of evidential justifications nor with the provision of evidential refutations. In the scientific domain, the mean difference in evidential justifications between high- and low-confidence items was 0.07, $t(115) = 1.97$, *ns*, and the mean difference in evidential refutations between high- and low-confidence items was −0.05, $t(70) = 1.80$, *ns*. In the supernatural domain, the corresponding differences were 0.06, $t(115) = 1.45$, *ns*, and −0.02, $t(70) = 0.61$, *ns*. In contrast to the lack of association between confidence ratings and evidential responses, there was a strong association between confidence ratings and *consensus estimates*. In both the scientific domain and the supernatural domain, the mean difference in consensus ratings between high- and low-confidence items was approximately 1.1 on a 7-point scale—a statistically robust difference in both cases: scientific domain, $t(115) = 12.58$, $p < .001$; supernatural domain, $t(71) = 8.18$, $p < .001$. Thus, regardless of domain, participants' confidence ratings appeared to be grounded in their perception of how widely those beliefs are shared but not their capacity for reasoning about those beliefs in relation to evidence. Put differently, the variance in participants' confidence ratings from one belief to the next was not associated with their ability to cite evidence relevant to those beliefs but was associated with their perception of the consensus surrounding those beliefs.

To provide a qualitative sense of these data, the confidence ratings for individual items were binned into eight categories determined by three binary dimensions: content domain (scientific vs. supernatural), justification type (evidential vs. deferential), and perceived consensus (low vs. high, defined as the difference between a rating of 1 through 4 and a rating of 5 through 7). The means for each category are displayed in Figure 4. This analysis not only confirms that participants' confidence ratings were more closely linked to their consensus estimates than to their means of justification but also demonstrates the influence of consensus estimates across domains. Although participants' confidence ratings for the scientific items were higher than their confidence ratings for the supernatural items at the same level of perceived consensus, their ratings for scientific items at the lowest level of perceived consensus were comparable to their ratings for supernatural items at the highest level of perceived consensus. In other words, their confidence in low-consensus scientific items, like evolution and black holes, was comparable to their confidence in high-consensus supernatural items, like God and souls. The difference in participants' confidence ratings between scientific and supernatural domains does not therefore appear to be categorical. The two domains overlap at a boundary defined by perceived consensus, with participants espousing as much confidence in widespread supernatural beliefs as in less widespread scientific beliefs (see also Table 2). For instance, confidence ratings for the

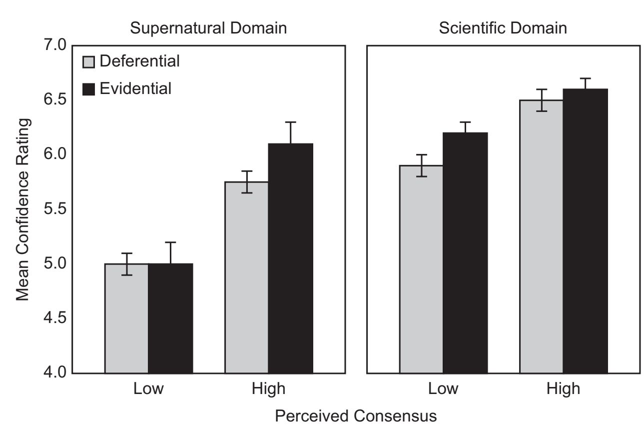


*Figure 4.* Mean confidence rating per item (+*SE*) as a function of domain (scientific, supernatural), justification type (evidential, nonevidential), and perceived consensus (low, high).

high-consensus supernatural item *God*, judged existent by 98 participants, were significantly lower than confidence ratings for the high-consensus scientific item *electrons*, $t(98) = 4.92$, $p < .001$, but were not significantly lower than confidence ratings for the low-consensus scientific item *black holes*, $t(98) = 0.36$, *ns*.

That said, it should be noted that consensus information is not the only contributor to participants' confidence, as confidence ratings for even the low-consensus supernatural items were high ($M = 5.0$ on a 7-point scale). It should also be noted that the direction of the relation between confidence ratings and consensus estimates is ambiguous. Participants' perceptions of the consensus surrounding a particular belief may have shaped their own confidence in that belief, or participants' confidence may have shaped their perceptions of consensus. Either way, the fact that confidence is more closely linked to perceptions of consensus than to means of justification suggests that participants' judgments were more heuristic than analytic (Evans, 2008; Stanovich & West, 1998).

## Discussion

Scientific entities like electrons and X-rays defy firsthand observation and everyday intuition, yet most people outside the scientific community still believe in their existence. Upon what kind of epistemological foundation do such beliefs rest? The present study explored this question by comparing students' scientific beliefs with their supernatural beliefs along four dimensions of epistemic import: personal confidence, perceived consensus, means of justification, and openness to revision. Although participants were significantly more confident in their scientific beliefs than in their supernatural beliefs, they were rarely able to identify evidence that might bear on the validity of those beliefs, either in the form of a justification or a potential refutation. This was particularly true of participants with (a) higher levels of supernatural belief and (b) lower levels of NOS understanding. Moreover, participants' confidence was related to their perception of how frequently other people would agree with their beliefs but not to their ability to cite evidential considerations relevant to those beliefs.

Three features of the data are particularly noteworthy. First, participants' modal form of justification was deference to the opinions and conclusions of others. That is, participants were more likely to reference the proximal source of their beliefs (i.e., the testimony of an accepted authority or the tenets of an accepted worldview) than to reference its distal source (i.e., reasons for accepting the testimony or tenets as true), both for scientific beliefs and supernatural beliefs. Although it could be argued that deference to "more knowledgeable others" is generally a rational course of action (Keil, 2010), this claim is undermined, in the present context, by a number of factors discussed below. One such factor was participants' insistence that many of their scientific beliefs were indefeasible—a view that runs counter to one of the major principles of science. Indeed, the majority of participants (55%) denied that anything could dissuade them of the existence of at least one scientific entity. This finding is consistent with Gottlieb's (2007) finding that 31% of adolescents attending secular schools and 66% attending religious schools claimed that their belief in God was infallible, but it extends Gottlieb's findings by demonstrating that convictions of infallibility are not restricted to the supernatural domain. Most students appear to be convinced that their beliefs about the existence of scientific entities are infallible as well.

Second, participants' confidence in their beliefs was more strongly associated with their perception of the consensus surrounding those beliefs than with their appreciation of the evidence in support of those beliefs. This finding complements a long tradition in social psychology demonstrating that consensus information tends to trump other forms of information, whether it be visual information (Asch, 1956; Bond & Smith, 1996), statistical information (Harries, Yaniv, & Harvey, 2004; Yaniv, Choshen-Hillel, & Milyavsky, 2009), or lexical information (Corriveau, Fusaro, & Harris, 2009; Fusaro & Harris, 2008). It also complements a long tradition in cognitive psychology demonstrating that people are generally blind to limitations in their own cognitive abilities, as manifested by overconfidence in the sufficiency of their knowledge (Arkes, Christensen, Lai, & Blumer, 1987; Koriat, Lichtenstein, & Fischhoff, 1980), overconfidence in the accuracy of their memory (Conway, Skitka, Hemmerich, & Kershaw, 2009; Talarico & Rubin, 2003), and overconfidence in the completeness of their understanding (Glenberg & Epstein, 1985; Rozenblit & Keil, 2002). Indeed, the lack of correlation between confidence ratings and evidential justifications in the present study is largely attributable to overconfidence: Participants who provided deferential reasons for their existence judgments were just as confident as those who provided evidential reasons, typically rating their confidence as a 6 out of 7 (85% confident) or a 7 out of 7 (100% confident), as shown in Figure 4.

Third, the findings obtained here with adults who had had multiple years of science instruction strongly mirror those obtained by Harris et al. (2006) with 5- to 6-year-old children, who had had virtually no science instruction. In Harris et al.'s study, children not only endorsed the existence of scientific entities more often than they endorsed the existence of supernatural entities but also reported (a) greater confidence in the existence of scientific entities and (b) greater levels of consensus surrounding the existence of scientific entities. They did not, however, provide different types of justifications for their endorsements. Instead, they tended to appeal to generalizations that presupposed the entity's existence in both cases (e.g., germs exist because "animals can have germs"; the Tooth Fairy exists because "she visits you when you lose a tooth"). Although it is unclear how the justification categories used in Harris et al.'s study relate to those used in the present study, it is telling that even young children exhibit greater sensitivity to the amount of consensus surrounding an extraordinary claim than to its evidential status. Although it would be unreasonable to expect 5- and 6-year-olds to provide evidence-based justifications, it is not unreasonable to expect college undergraduates to do so; yet, the data reported here suggests that, for many students, multiple years of science education does not substantially alter early developing patterns in the endorsement and justification of scientific beliefs.

As a whole, these findings imply that students' understanding of science as a body of knowledge is similar to their understanding of science as a method of inquiry, or lack thereof (Carey et al., 1989; Lederman et al., 2002; Schauble et al., 1995; Smith et al., 2000). Just as students conceive of science as problem solving rather than inquiry, they justify their scientific beliefs with appeals to intuition and authority rather than evidence. And just as students think that scientists are in the business of "proving their ideas true," they think that certain scientific entities have been proved to exist beyond a shadow of a doubt. These findings not only complement existing findings on students' scientific epistemologies but also point to the possibility that misconceptions about the process of science may actually *give rise to* misconceptions about the products of science. The products of science are not, after all, well differentiated from the products of magical or religious modes of thought.

This similarity in scientific and supernatural beliefs actually highlights a potentially important instructional implication: that supernatural beliefs should not be quarantined from scientific ones in the classroom but should stand subject to the same kinds of empirical and theoretical scrutiny. To quarantine them assumes that students understand that supernatural claims rest on qualitatively different grounds than scientific claims, yet many students clearly do not. In fact, the findings reviewed here suggest that the decision to exclude supernatural beliefs from the classroom may strike many students as a matter of convention rather than as a matter of principle, and by doing so educators are actually foregoing a natural, and potentially influential, inroad into students' epistemologies of science. The negative implication of such an instructional technique is, of course, raising the idea that students' supernatural beliefs are not on par with their scientific beliefs, but the present findings suggest that, without proper guidance, students will not develop an appreciation of this idea on their own.

## Consideration of Alternative Interpretations

The foregoing discussion has assumed that participants' reliance on nonevidential considerations when justifying and evaluating their scientific beliefs should be considered epistemologically problematic, but is it? One way of interpreting the predominance of such considerations is that they are actually an artifact of the study's methodology. Perhaps participants were not sufficiently engaged with the task, or they were sufficiently engaged but not in the intended manner. For instance, some participants may have interpreted questions like "Why do you believe in the existence of evolution?" not as a query for *evidence* but as a query for *reasons*, with the relevant evidence (e.g., fossils, anatomical homologies, genetic homologies) implied by the cited reasons (e.g., that evolution is included in science textbooks). Not citing evidence in support of one's beliefs is not the same as failing to value evidence as a means of justification (see Brem & Rips, 2000; Sandoval & Cam, 2011), and participants may simply have been unaware of the expectation that they cite evidence.

Clearer task demands may thus have yielded more evidential responding, but there are at least three reasons to doubt that the present rates of evidential responding are *entirely* artifactual. First, the amount of information participants needed to provide in order for their responses to be coded as "evidential" was minimal, particularly in light of the participants' educational backgrounds. It would have been unreasonable to expect participants to recite detailed accounts of the evidence for X-rays or evolution years after having learned it, and thus responses were coded as "evidential" so long as they referenced a hint of that evidence. In fact, justifications as simple as "learning about Marie Curie and her studies" (for X-rays) and "it explains the biology of our planet better than anything else" (for evolution) were coded as evidential, even though the evidence contained in such justifications is largely implicit. Deferential justifications, however, referenced no evidential content whatsoever (e.g., "scientists have proved that it exists," "it was in my chemistry textbook"), and it seems unlikely that participants who provided such justifications as their modal justification type for both scientific and supernatural beliefs would provide substantively different justifications in a different context.

Second, participants answered questions not only about entities they believed to exist but also about entities they believed not to exist, which created a motivation to differentiate the two in as substantive a way as possible. A participant who claimed to believe in genes but not souls, for example, was asked to justify both beliefs on the same questionnaire and might thus have strived to identify considerations that would legitimately differentiate the two, lest she appear unprincipled in her beliefs. Although it is debatable whether participants actually concerned themselves with these kinds of self-presentation demands, it is unlikely that participants' failure to provide evidential justifications stemmed wholly from a lack of motivation. Indeed, multiple participants revealed, during the poststudy debriefing, that their participation in the study caused them to realize how groundless many of their scientific beliefs actually were.

Third, interpreting participants' lack of evidential responding as a performance error, rather than as a genuine reflection of their epistemological commitments, is less plausible when those responses are considered in light of the entire nexus of results; namely, (a) deferential responding was associated with lower levels of NOS understanding, (b) deferential responding was associated with higher levels of supernatural belief, and (c) participants who provided deferential justifications for their scientific beliefs tended to provide justifications *of the same form* for their supernatural beliefs. This last finding is particularly telling, for although it may be reasonable to assume that a participant who believes in electrons because he has "always been taught they exist" understands that deference to authority is justified in this circumstance, it is less reasonable to make that assumption when (a) he provides the same justification for why he believes in angels and (b) he claims to be 100% confident of the existence of both.

In short, the requirements for providing evidential justifications were relatively low, and the motivation for doing so was relatively high, yet participants still tended to provide deferential justifications for their scientific beliefs that mirrored those they provided for their supernatural beliefs. That said, one might still argue that deferential responses are no less epistemologically sound than evidential ones. After all, deference to experts is common even among professional scientists, as the cognitive labor entailed by modern-day scientific inquiry is far too great for any one discipline, let alone any one individual (Kitcher, 2001). The labor must be divided—a fact appreciated by both scientists and nonscientists alike. Indeed, recent research by Keil and colleagues has shown that even preschool-aged children are sensitive to the implications of this division, demonstrating a rudimentary appreciation of how knowledge is clustered in other minds (Keil, Stein, Webb, Billings, & Rozenblit, 2008), what kinds of principles underlie expert knowledge (Keil, 2010), and who to query for the answers to particular empirical questions (Lutz & Keil, 2002). Do participants' deferential responses merely reflect an early developing—and empirically justifiable—appreciation of the division of cognitive labor?

Certainly, the *predominance* of deferential responses is well explained by an appreciation of the division of cognitive labor, but the *adequacy* of such responses is not. What these responses appear to belie is not merely a shallow epistemology but an ill-founded epistemology, that is, an epistemology in which all claims, scientific or supernatural, are on equal footing and the decision of whether or not to endorse such claims is based either on personal predilection or on the frequency with which others tend to endorse them (see Kuhn, 1991, for a similar account in the domain of argumentation). As noted above, deferential justifications for belief in the existence of scientific entities did not occur in a vacuum; rather, they co-occurred with higher levels of supernatural belief and lower levels of NOS understanding. Furthermore, not all participants justified their scientific beliefs by deference. Some actually cited evidence, and those participants (roughly 20% of the sample) tended to show other differences in epistemological understanding, namely, (a) an increased ability to identify counterevidence to their scientific beliefs and (b) an increased tendency to cite evidence even in support of their supernatural beliefs. Individual differences of this nature imply that deferential responses are not an inevitable byproduct of the division of cognitive labor (and one's sensitivity to it) but are instead a particular manifestation of how one has construed the epistemic status of scientific claims.

That said, additional research is needed to determine whether, and how, the response patterns elicited here vary by such factors as age, education, and context. If, as argued here, such patterns represent the workings of an ill-founded epistemology, then they should be robust to changes in how belief justifications and refutations are elicited or measured. They may even be robust to instructional interventions regarding the evidential foundations of scientific claims. Indeed, one fruitful line of inquiry in future research would be to compare students' response patterns across different types of interventions, for example, interventions that focus on the difference between evidential and deferential modes of argumentation versus interventions that focus on the difference between scientific claims and supernatural claims. Both would likely increase students' evidential responding, but one might prove more effective than the other or one might prove more comprehensive than the other, influencing aspects of epistemological understanding not explicitly covered by the intervention (e.g., confidence ratings, NOS understanding). Such findings would be useful not only for pedagogical purposes but also for determining which facets of epistemological understanding are most central to the phenomena assessed here.

## Relations to Other Measures of Epistemological Understanding

The present study outlines a new approach to studying students' epistemological commitments. How this approach compares with previous approaches is not immediately obvious, for some aspects of epistemological understanding may be more foundational than others and some means of assessing epistemological understanding may be more ecologically valid than others. Below, I conclude by considering how students' understanding of the evidential foundations of scientific claims relates to three other aspects of epistemological understanding, with an emphasis on identifying directions for future research.

First, what is the relation between students' beliefs about the nature of science, in general, and their ability to justify and evaluate particular scientific claims? Empirically, these two sets of beliefs appear to be correlated, with higher levels of NOS understanding associated with higher levels of evidential responding and lower levels of deferential responding (as shown in Figure 3), but what connects the two conceptually? One possibility, noted above, is that misconceptions about the process of science give rise to misconceptions about the products of science. Consistent with this possibility, participants' judgments and justifications for six such products—the concepts *black hole*, *electron*, *evolution*, *fluoride*, *gene*, and *X-ray*—revealed seemingly pervasive misconceptions about science as a process, including (a) that proponents of scientific claims stand on relatively equal footing with the proponents of supernatural claims, (b) that the validity of scientific claims can be accurately gauged by popular consensus, and (c) that certain scientific entities have been proven to exist beyond a shadow of a doubt. Nevertheless, further work needs to be done to clarify this relation, particularly in the context of more ecologically valid measures of NOS understanding. One such approach would be to administer the belief-justification task reported here to individuals with varying levels of expertise in the actual practice of science (e.g., high school students, undergraduate science majors, graduate students in science, professional scientists) on the prediction that expertise in science changes the way one justifies and evaluates one's scientific beliefs, even those outside one's area of expertise.

Second, what is the relation between students' beliefs about the *nature of knowledge* and their ability to justify or evaluate particular scientific claims? Again, participants' performance suggests they hold a number of misconceptions in this domain, including (a) that claims supported by multiple forms of inquiry and multiple sources of evidence (i.e., scientific claims) stand on relatively equal footing to those that do not (i.e., supernatural claims), (b) that appeals to authority are as valuable as appeals to evidence, and (c) that empirical claims can be "elevated" to a point at which empirical evidence no longer bears on their validity. These misconceptions seem to be characteristic of what has been termed a "relativist" or "multiplist" conception of knowledge (Hofer & Pintrich, 1997; Kuhn, Cheney, & Weinstock, 2000); yet, it is unclear whether the individuals who implicitly conveyed such misconceptions in their justifications and refutations would have been diagnosed as relativists using more standard measures of epistemological understanding. Still, one of the most intriguing findings to emerge from the literature on naïve epistemology is that levels of epistemological understanding predict evidence-based reasoning among jurors in a court of law (Kuhn, Weinstock, & Flaton, 1994; Weinstock & Cronin, 2003), implying that this construct has high predictive validity for reasoning about content-rich problems, including, perhaps, the type of reasoning assessed here.

Finally, what is the relation between domain-specific knowledge and the ability to justify and evaluate domain-relevant claims? Numerous studies have documented robust individual differences in the understanding of domain-specific concepts (e.g., McCloskey, 1983; Shtulman, 2006; Slaughter & Lyons, 2003; Slotta & Chi, 2006; Smith, 2007), and it seems plausible that those who hold an accurate understanding of such concepts are more likely to understand (a) the data that bear on those concepts, (b) the methods that shape those concepts, and (c) the theories that depend on those concepts. Consider, for instance, how understanding evolution might facilitate an appreciation of the evidence in support of evolution. Those who possess such an understanding would be more likely than those who do not to appreciate (a) how phenomena like the distribution of fossils in the Earth's crust or the distribution of genes across different genomes bear on the validity of evolutionary claims, (b) how methods like comparative morphology and comparative genetics give rise to new insights about the nature of evolution, and (c) how concepts like evolution and natural selection serve to unify such disparate phenomena as mimicry in butterflies and bacterial resistance to antibiotics (see Shtulman & Calabi, 2012, for more discussion of these issues).

If students' failure to provide evidential justifications and evidential refutations for their scientific beliefs is a product of poor understanding (or misunderstanding) of the relevant concepts, then interventions that prove effective at increasing students' understanding should also prove effective at increasing their evidential reasoning. Alternatively, interventions that prove effective at increasing students' evidential reasoning may also prove effective at increasing their understanding. Whether or not both effects could be achieved with the same intervention, the relation between conceptual understanding and evidential reasoning deserves additional attention. The ability to cite evidence in support of one's beliefs is, after all, a primary form of scientific literacy and one that needs to be in place if citizens are to make informed decision about public policies of a scientific nature, like whether genetically modified foods should be banned from grocery stores, whether fluoride should be added to tap water, or whether stem cells should be made available for research on cell differentiation. The findings documented here suggest that this ability is lacking even among individuals who have received multiple years of science education. Further research on this particular form of epistemological understanding thus promises to shed light on a problem of both practical and theoretical importance.

## References


American Association for the Advancement of Science. (2002). *AAAS board resolution on intelligent design theory*. Washington, DC: Author.

Arkes, H. R., Christensen, C., Lai, C., & Blumer, C. (1987). Two methods of reducing overconfidence. *Organizational Behavior and Human Decision Processes, 39,* 133–144. doi:10.1016/0749-5978(87)90049-5

Asch, S. E. (1956). Studies of independence and conformity: A minority of one against a unanimous majority. *Psychological Monographs*, *70*.

Au, T. K. F. (1994). Developing an intuitive understanding of substance kinds. *Cognitive Psychology, 27,* 71–111. doi:10.1006/cogp.1994.1012

Baxter Magolda, M. B. (1992). *Knowing and reasoning in college: Gender-related patterns in students' intellectual development*. San Francisco, CA: Jossey-Bass.

Blancke, S., De Smedt, J., De Cruz, H., Boudry, M., & Braeckman, J. (2011). The implications of the cognitive sciences for the relation between religion and science education: The case of evolutionary theory. *Science & Education, 21,* 1167–1184.

Bond, R., & Smith, P. B. (1996). Culture and conformity: A meta-analysis of studies using Asch's (1952b, 1956) line judgment task. *Psychological Bulletin, 119,* 111–137. doi:10.1037/0033-2909.119.1.111

Brem, S. K., Ranney, M., & Schindel, J. (2003). Perceived consequences of evolution: College students perceive negative personal and social impact in evolutionary theory. *Science Education, 87,* 181–206. doi:10.1002/sce.10105

Brem, S. K., & Rips, L. J. (2000). Explanation and evidence in informal argument. *Cognitive Science, 24,* 573–604. doi:10.1207/s15516709cog2404_2

Brockman, J. (2006). *Intelligent thought: Science versus the intelligent design movement*. New York, NY: Vintage Books.

Carey, S. (2009). *The origin of concepts*. New York, NY: Oxford University Press. doi:10.1093/acprof:oso/9780195367638.001.0001

Carey, S., Evans, R., Honda, M., Jay, E., & Unger, C. (1989). An experiment is when you try and see if it works: A study of junior high school students' understanding of the construction of scientific knowledge. *International Journal of Science Education, 11,* 514–529. doi:10.1080/0950069890110504

Chinn, C. A., & Brewer, W. F. (1998). An empirical test of a taxonomy of responses to anomalous data in science. *Journal of Research in Science Teaching, 35,* 623–654. doi:10.1002/(SICI)1098-2736(199808)35:6<623::AID-TEA3>3.0.CO;2-O

Chinn, C. A., & Brewer, W. F. (2001). Models of data: A theory of how people evaluate data. *Cognition and Instruction, 19,* 323–393. doi:10.1207/S1532690XCI1903_3

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Ealbaum.

Conway, A. R. A., Skitka, L. J., Hemmerich, J. A., & Kershaw, T. C. (2009). Flashbulb memory for 11 September 2001. *Applied Cognitive Psychology, 23,* 605–623. doi:10.1002/acp.1497

Corriveau, K. H., Fusaro, M., & Harris, P. L. (2009). Going with the flow: Preschoolers prefer nondissenters as informants. *Psychological Science, 20,* 372–377. doi:10.1111/j.1467-9280.2009.02291.x

Dawkins, R. (2009). *The greatest show on earth: The evidence for evolution*. New York, NY: Free Press.

Evans, J. St. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology, 59,* 255–278. doi:10.1146/annurev.psych.59.103006.093629

Frazer, J. G. (1998). *The Golden Bough*. Oxford, UK: Oxford University Press. (Original work published 1922)

Fusaro, M., & Harris, P. L. (2008). Children assess informant reliability using bystanders' nonverbal cues. *Developmental Science, 11,* 771–777. doi:10.1111/j.1467-7687.2008.00728.x

Glenberg, A. M., & Epstein, W. (1985). Calibration of comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 702–718. doi:10.1037/0278-7393.11.1-4.702

Gottlieb, E. (2007). Learning how to believe: Epistemic development in cultural context. *The Journal of the Learning Sciences, 16,* 5–35.

Harries, C., Yaniv, I., & Harvey, N. (2004). Combining advice: The weight of a dissenting opinion in the consensus. *Journal of Behavioral Decision Making, 17,* 333–348. doi:10.1002/bdm.474

Harris, P. L., & Koenig, M. A. (2006). Trust in testimony: How children learn about science and religion. *Child Development, 77,* 505–524. doi:10.1111/j.1467-8624.2006.00886.x

Harris, P. L., Pasquini, E. S., Duke, S., Asscher, J., & Pons, F. (2006). Germs and angels: The role of testimony in young children's ontology. *Developmental Science, 9,* 76–96. doi:10.1111/j.1467-7687.2005.00465.x

Hofer, B. K., & Pintrich, P. (1997). The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning. *Review of Educational Research, 67,* 88–140.

Inter-Academy Panel. (2006). *IAP statement on the teaching of evolution*. Trieste, Italy: Author.

Keil, F. C. (2010). The feasibility of folk science. *Cognitive Science, 34,* 826–862. doi:10.1111/j.1551-6709.2010.01108.x

Keil, F. C., Stein, C., Webb, L., Billings, V. D., & Rozenblit, L. (2008). Discerning the division of cognitive labor: An emerging understanding of how knowledge is clustered in other minds. *Cognitive Science, 32,* 259–300. doi:10.1080/03640210701863339

King, P. M., & Kitchener, K. S. (1994). *Developing reflective judgment: Understanding and promoting intellectual growth and critical thinking in adolescents and adults*. San Francisco, CA: Jossey-Bass.

Kitcher, P. (2001). *Science, truth, and democracy*. Oxford, UK: Oxford University Press. doi:10.1093/0195145836.001.0001

Klahr, D., Fay, A. L., & Dunbar, K. (1993). Heuristics for scientific experimentation: A developmental study. *Cognitive Psychology, 25,* 111–146. doi:10.1006/cogp.1993.1003

Koehler, J. J. (1993). The influence of prior beliefs on scientific judgments of evidence quality. *Organizational Behavior and Human Decision Processes, 56,* 28–55. doi:10.1006/obhd.1993.1044

Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 107–118. doi:10.1037/0278-7393.6.2.107

Koslowski, B., Marasia, J., Chelenza, M., & Dublin, R. (2008). Information becomes evidence when an explanation can incorporate it into a causal framework. *Cognitive Development, 23,* 472–487. doi:10.1016/j.cogdev.2008.09.007

Kuhn, D. (1991). *The skills of argument*. New York, NY: Cambridge University Press. doi:10.1017/CBO9780511571350

Kuhn, D., Amsel, E., & O'Loughlin, M. (1988). *The development of scientific thinking skills*. Orlando, FL: Academic Press.

Kuhn, D., Cheney, R., & Weinstock, M. (2000). The development of epistemological understanding. *Cognitive Development, 15,* 309–328. doi:10.1016/S0885-2014(00)00030-7

Kuhn, D., & Udell, W. (2003). The development of argument skills. *Child Development, 74,* 1245–1260. doi:10.1111/1467-8624.00605

Kuhn, D., Weinstock, M., & Flaton, R. (1994). How well do jurors' reason? Competence dimensions of individual variation in a juror reasoning task. *Psychological Science, 5,* 289–296. doi:10.1111/j.1467-9280.1994.tb00628.x

Lederman, N. G., Abd-El-Khalick, F., Bell, R. L., & Schwartz, R. S. (2002). Views of nature of science questionnaire: Toward valid and meaningful assessment of learners' conceptions of nature of science. *Journal or Research in Science Teaching, 39,* 497–521. doi:10.1002/tea.10034

Lederman, N. G., & O'Malley, M. (1990). Students' perceptions of tentativeness in science: Development, use, and sources of change. *Science Education, 74,* 225–239. doi:10.1002/sce.3730740207

Legare, C. H., Evans, E. M., Rosengren, K. S., & Harris, P. L. (2012). The coexistence of natural and supernatural explanations across cultures and development. *Child Development, 83,* 779–793. doi:10.1111/j.1467-8624.2012.01743.x

Leiserowitz, A. (2005). American risk perceptions: Is climate change dangerous? *Risk Analysis, 25,* 1433–1442. doi:10.1111/j.1540-6261.2005.00690.x

Leiserowitz, A. (2006). Climate change risk perception and public policy preferences: The role of affect, imagery, and values. *Climate Change, 77,* 45–72. doi:10.1007/s10584-006-9059-9

Lindeman, M., Svedholm, A. M., Takada, M., Lonnqvist, J. E., & Verkasalo, M. (2011). Core knowledge confusions among university students. *Science & Education, 20,* 439–451. doi:10.1007/s11191-009-9210-x

Lombrozo, T., Shtulman, A., & Weisberg, M. (2006). The intelligent design controversy: Lessons from psychology and education. *Trends in Cognitive Sciences, 10,* 56–57. doi:10.1016/j.tics.2005.12.001

Lombrozo, T., Thanukos, A., & Weisberg, M. (2008). The importance of understanding the nature of science for accepting evolution. *Evolution: Education & Outreach, 1,* 290–298. doi:10.1007/s12052-008-0061-8

Lutz, D. J., & Keil, F. C. (2002). Early understandings of the division of cognitive labor. *Child Development, 73,* 1073–1084. doi:10.1111/1467-8624.00458

McCauley, R. N. (2000). The naturalness of religion and the unnaturalness of science. In F. Keil & R. Wilson (Eds.), *Explanation and cognition* (pp. 61–85). Cambridge, MA: MIT Press.

McCloskey, M. (1983). Naïve theories of motion. In D. Gentner & A. Stevens (Eds.), *Mental models* (pp. 299–324). Hillsdale, NJ: Erlbaum.

Moore, D. W. (2005). *Three in four Americans believe in paranormal*. Princeton, NJ: The Gallup Organization.

Nakhleh, M. B., & Samarapungavan, A. (1999). Elementary school children's beliefs about matter. *Journal of Research in Science Teaching, 36,* 777–805. doi:10.1002/(SICI)1098-2736(199909)36:7<777::AID-TEA4>3.0.CO;2-Z

National Academy of Sciences. (1999). *Science and Creationism: A view from the National Academy of Sciences* (2nd ed.). Washington, DC: Author.

Novick, L. R., Shade, C. K., & Catley, K. M. (2010). Linear versus branching depictions of evolutionary history: Implications for diagram design. *Topics in Cognitive Science, 1,* 1–24.

Rozenblit, L., & Keil, F. (2002). The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive Science, 26,* 521–562. doi:10.1207/s15516709cog2605_1

Sá, W. C., Kelley, C. N., Ho, C., & Stanovich, K. E. (2005). Thinking about personal theories: Individual differences in the coordination of theory and data. *Personality and Individual Differences, 38,* 1149–1161. doi:10.1016/j.paid.2004.07.012

Sandoval, W. A., & Cam, A. (2011). Elementary children's judgments of the epistemic status of sources of justification. *Science Education, 95,* 383–408. doi:10.1002/sce.20426

Schauble, L., Glaser, R., Duschl, R. A., Schulze, S., & John, J. (1995). Students' understanding of the objectives and procedures of experimentation in the science classroom. *Journal of the Learning Sciences, 4,* 131–166. doi:10.1207/s15327809jls0402_1

Schauble, L., Klopfer, L. E., & Raghavan, K. (1991). Students' transition from an engineering model to a science model of experimentation. *Journal of Research in Science Teaching, 28,* 859–882. doi:10.1002/tea.3660280910

Shtulman, A. (2006). Qualitative differences between naive and scientific theories of evolution. *Cognitive Psychology, 52,* 170–194. doi:10.1016/j.cogpsych.2005.10.001

Shtulman, A., & Calabi, P. (2012). Cognitive constraints on the understanding and acceptance of evolution. In K. S. Rosengren, S. Brem, E. M. Evans, & G. Sinatra (Eds.), *Evolution challenges: Integrating research and practice in teaching and learning about evolution* (pp. 47–65). Cambridge, UK: Oxford University Press.

Shtulman, A., & Schulz, L. (2008). The relation between essentialist beliefs and evolutionary reasoning. *Cognitive Science, 32,* 1049–1062. doi:10.1080/03640210801897864

Slaughter, V., & Lyons, M. (2003). Learning about life and death in early childhood. *Cognitive Psychology, 46,* 1–30. doi:10.1016/S0010-0285(02)00504-2

Slotta, J. D., & Chi, M. T. H. (2006). Helping students understand challenging topics in science through ontology training. *Cognition and Instruction, 24,* 261–289. doi:10.1207/s1532690xci2402_3

Smith, C. L. (2007). Bootstrapping processes in the development of students' commonsense matter theories: Using analogical mappings, thought experiments, and learning to measure to promote conceptual restructuring. *Cognition and Instruction, 25,* 337–398. doi:10.1080/07370000701632363

Smith, C. L., Maclin, D., Houghton, C., & Hennessey, M. G. (2000). Sixth-grade students' epistemologies of science: The impact of school science experiences on epistemological development. *Cognition and Instruction, 18,* 349–422. doi:10.1207/S1532690XCI1803_3

Smith, C. L., & Wenk, L. (2006). Relations among three aspects of first-year college students' epistemologies of science. *Journal of Research in Science Teaching, 43,* 747–785. doi:10.1002/tea.20113

Songer, N. B., & Linn, N. B. (1991). How do students views of science influence knowledge integration? *Journal of Research in Science Teaching, 28,* 761–784. doi:10.1002/tea.3660280905

Stanovich, K. E., & West, R. F. (1998). Individual differences in rational thought. *Journal of Experimental Psychology: General, 127,* 161–188. doi:10.1037/0096-3445.127.2.161

Talarico, J. M., & Rubin, D. C. (2003). Confidence, not consistency, characterizes flashbulb memories. *Psychological Science, 14,* 455–461. doi:10.1111/1467-9280.02453

Toulmin, S. (1964). *The uses of argument*. New York, NY: Cambridge University Press.

Vosniadou, S. (1994). Capturing and modeling the process of conceptual change. *Learning and Instruction, 4,* 45–69. doi:10.1016/0959-4752(94)90018-3

Weinstock, M., & Cronin, M. A. (2003). The everyday production of knowledge: Individual differences in epistemological understanding and juror-reasoning skill. *Applied Cognitive Psychology, 17,* 161–181. doi:10.1002/acp.860

Winseman, A. L. (2004). *Eternal destinations: Americans believe in Heaven, Hell*. Princeton, NJ: The Gallup Organization.

Yaniv, I., Choshen-Hillel, S., & Milyavsky, M. (2009). Spurious consensus and opinion revision: Why might people be more confident in their less accurate judgments? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 558–563. doi:10.1037/a0014589

## Appendix

## The Nature of Science (NOS) Assessment

1. Which of the following best describes the purpose of an experiment?

A. To make a discovery
B. To test an idea*
C. To find a solution
D. To prove a hypothesis true

2. Which of the following most influences a scientist's decision as to what experiments to conduct?

A. Public interest
B. Personal interest
C. Previous research in the field*
D. Potential application of the findings

3. Which of the following is most important to the practice of science in general?

A. The use of instruments
B. The use of measurement
C. The use of a systematic protocol
D. The collection of data*

4. Which of the following is generally most valued by the scientific community?

A. Facts
B. Hypotheses
C. Laws
D. Theories*

5. Which of the following best describes the nature of a scientific theory?

A. A well-supported explanation*
B. A well-educated guess
C. A well-documented finding
D. An indisputable argument

6. How common is it for scientists who study the same topic to disagree?

A. Very common*
B. Moderately common
C. Moderately uncommon
D. Very uncommon

7. Disagreements in science most typically arise for which of the following reasons?

A. Different scientists gather different kinds of data
B. Different scientists interpret the same data in different ways*
C. No one is ever 100% correct
D. No question has a single right answer

Correct responses are indicated with an asterisk.

Received May 2, 2011
Revision received August 23, 2012
Accepted September 4, 2012 ■