Evaluative Effects on Knowledge Attributions¹

James R. Beebe (University at Buffalo)

Forthcoming in Justin Sytsma and Wesley Buckwalter (eds.),

*The Blackwell Companion to Experimental Philosophy*

Experimental philosophers have investigated various ways in which non-epistemic evaluations can affect knowledge attributions. For example, several teams of researchers (Beebe & Buckwalter 2010; Beebe & Jensen 2012, Schaffer & Knobe 2012, Beebe & Shea 2013, Buckwalter 2014b, Turri 2014) report that the goodness or badness of an agent’s action can affect whether the agent is taken to have certain kinds of knowledge. These findings raise important questions about how patterns of folk knowledge attributions should influence philosophical theorizing about knowledge.

The contemporary starting point for philosophical discussions of the nature of knowledge is the justified true belief (or JTB) model. According to this traditional perspective, in order to know something, e.g., that water is H₂O, it must be true that water is H₂O. This requirement of truth (also known as factivity) specifies that, no matter how firmly you might believe something or how strong your reasons for believing it might be, if it is not true, you cannot know it. In addition to truth, there are also requirements regarding belief and justification. If it is true that water is H₂O, but you have no reason for thinking that it is, you cannot be said to know it. Even if you could bring yourself to believe that water is H₂O without having any reason for thinking that it is, and it turned out that your belief was correct, this would not be enough for knowledge. Lucky guesses do not count as knowledge. Furthermore, even if you had reason for thinking that

¹ The title for this chapter is inspired by remarks from Turri (2014, 122).
water is H₂O and if the truth requirement were satisfied, if you did not in fact believe that water is H₂O, it would not be true that you know that water is H₂O. Thus, according to the JTB model, each of the justification, truth, and belief conditions is necessary for knowledge, and together they are sufficient.

As is well-known, the JTB model first came under significant philosophical attack by Edmund Gettier (1963), who used cleverly crafted thought experiments to argue that justification, truth, and belief are not in fact sufficient for knowledge.² More recently, mainstream and experimental philosophers have begun to question whether the various components of the JTB model are even necessary. Allan Hazlett (2010; 2012), for example, has suggested that knowledge without truth might be possible.³ Blake Myers-Schulz and Eric Schwitzgebel (2013), Dylan Murray, Justin Sytsma, and Jonathan Livengood (2013), and Beebe (2013) provide experimental evidence suggesting that—at least in the minds of many non-philosophers—belief is not always required for knowledge. Finally, Crispin Sartwell (1991; 1992) and David Sackris and Beebe (2014) have explored the possibility that justification may not be necessary.

On another front, defenders of ‘pragmatic encroachment’ in epistemology (e.g., Fantl & McGrath, 2002, 2007, 2009; Hawthorne, 2004; Stanley, 2005; Hawthorne & Stanley, 2008) have recently argued that, in addition to justification, truth, and belief, how much is at stake for a believing subject can also determine whether the subject has knowledge. If two people have the same evidence for the same belief and that belief has the same truth value in both cases, epistemological tradition dictates that both beliefs must either count or fail to count as knowledge. Proponents of pragmatic encroachment, however, reject the idea that purely

² Cf. Turri (2015) for an overview of work in experimental philosophy on intuitions in Gettier cases.
³ However, cf. Buckwalter (2014a) for an empirical investigation of Hazlett’s claims.
epistemic factors like evidence or reliability are the only things that distinguish knowledge from mere true belief. If the practical costs of being wrong are high for one subject but low for another, the former may fail to have knowledge while the latter succeeds—even if everything else is held constant. Various experimental philosophers (e.g., Buckwalter 2010, Feltz & Zarpentine 2010, May et al. 2010, Pinillos 2012, Pinillos & Simpson 2014, Sripada & Stanley 2012, Phelan forthcoming, Buckwalter & Schaffer forthcoming) have made important contributions to the debate about the extent to which ordinary practices of knowledge attribution allow for or include something like pragmatic encroachment. It is important to note that the question at hand is not the trivial one of whether stakes can influence knowledge attributions in the way that non-epistemic factors like being drunk, tired, racist, or distracted can. Rather, the issue is whether a non-epistemic factor like stakes can influence knowledge attributions in a way that is proper.

It is against this background of debate over possible non-epistemic conditions on knowledge that evaluative effects on knowledge attributions are best understood. The epistemic side-effect effect (or ESEE) is the finding that factors like the goodness or badness of the outcome of an agent’s action or facts about whether the agent has fulfilled or violated a relevant social norm can affect whether that agent is taken to have certain kinds of knowledge (Beebe & Buckwalter 2010; Beebe & Jensen 2012, Schaffer & Knobe 2012, Beebe & Shea 2013, Buckwalter 2014b, Turri 2014). The descriptive or empirical question is what psychological factors are driving this effect. The normative, philosophical question is whether we should view these factors as biases or as reflecting folk competence with the concept of knowledge.

Work on the ESEE was originally inspired by Joshua Knobe’s (2003a, 2003b, 2004; 2010) important discovery of the ‘Knobe effect’ or ‘side-effect effect’—probably the most
famous result in experimental philosophy. Knobe found that individuals are more likely to say that a bad side-effect was brought about intentionally than a good one.\footnote{Cf. Cova (2015) for an overview of empirical work on this issue.} He began by presenting participants with either the help or the harm version of the following vignette:

> The vice-president of a company went to the chairman of the board and said, "We are thinking of starting a new program. It will help us increase profits, but it will also help/harm the environment." The chairman of the board answered, "I don’t care at all about helping/harming the environment. I just want to make as much profit as I can. Let’s start the new program." They started the new program. Sure enough, the environment was harmed.

Knobe then asked participants whether the chairman brought about the side-effect intentionally. In the harm condition, 83% of participants thought he harmed the environment intentionally, but only 23% thought he helped the environment intentionally. Knobe’s result challenged the traditional view that assessments of intentionality are made independently of knowledge of the consequences of an agent’s action.

Wesley Buckwalter wondered whether the same kind of asymmetrical attribution of a folk psychological attitude might be found if participants were asked about the chairman’s knowledge. He asked several hundred undergraduates whether the chairman knew that the environment would be helped or harmed, instructing them to register their opinions on a seven-point scale with –3 labeled as ‘the chairman didn’t know,’ and 3 labeled as ‘the chairman knew.’ Buckwalter found that participants were more inclined to think the chairman had knowledge in the harm condition ($M = 2.25$) than in the help condition ($M = 0.91$) (Beebe & Buckwalter 2010). Almost twice as many participants chose the strongest possible affirmation of the chairman’s knowledge (viz., response ‘3’) in the harm case (67.5%) as in the help case (35.5%). Beebe and
Jensen (2012) extended this line of research by finding the ESEE in cases where the central protagonist (a) implements a corporate restructuring plan that either increases or decreases sales in New Jersey for the next quarter, (b) implements a new profit-increasing policy at a movie studio that makes movies either better or worse from an artistic standpoint, or (c) makes organizational changes at a corporation that either fulfills or violates a racial identification law in Nazi Germany. In other words, Beebe and Jensen found that individuals were more likely to attribute knowledge to a protagonist whose action was bad in some respect than to a nearly equivalent protagonist whose action was good or neutral.

The ESEE has important consequences for understanding the Knobe effect because most of the scholars who have proposed explanations of the effect base their explanations upon the assumption that the chairman knows that the side-effects in the help and the harm conditions will occur. Hugh McCann (2005, 739), for example, writes, “The situation of Knobe’s chairman, however, is not one of negligence: in both vignettes, he knows perfectly well what he is doing.” Many other scholars have followed suit, viewing the side-effects as being “anticipated” or “foreseen.”5 But if participants are not equally inclined to attribute knowledge to the chairman in the contrasting conditions, explanations of why they asymmetrically attribute intentionality that are based upon this erroneous assumption will likely be incorrect.

The ESEE also has the potential to change our understanding of the nature of knowledge. If individuals are not making an error when they allow the goodness or badness of an action undertaken in light of the belief that \( p \) to affect their judgments about whether someone knows that \( p \), then epistemologists will have one more non-epistemic factor—in addition to stakes—to

---

consider incorporating into their analyses of knowledge. However, several philosophical and empirical challenges must be answered before such a move can be fully warranted.

Jonathan Weinberg (2014) has suggested that the size of the ESEE is “not terribly dramatic, and thus would be better evidence that the Knobe effect is here a source of noise, not signal, concerning what knowledge really is or isn’t.” However, the effect observed by Beebe and Buckwalter (2010) was medium in size ($r = .35$), and when Beebe and Jensen (2012) reran Beebe and Buckwalter’s experiment using a forced-choice (‘the chairman knows’ vs. ‘the chairman doesn’t know’) answer format and a Likert scale that ranged from 1 to 7 instead of from -3 to 3, they observed large effect sizes in both cases ($r’s = .51$ and .7). While these findings raise important questions about what the effect size for the ESEE really is, the size of the observed effects strongly suggests that the ESEE is more than mere noise.

Another question that has arisen in the discussion of the ESEE is whether the effect might result from the fact that it is often easier to make things worse than it is to make things better. If, for example, the probability of succeeding in harming the environment is higher than the probability of succeeding in helping it, it would be easier for the chairman to know that his actions will harm the environment than for him to know that he will help it. Thus, standard presentations of the chairman and the environment case may err in assuming that the chairman’s epistemic position with respect to the belief in question is equally strong in both conditions. If this is correct, the ESEE will have nothing of substance to teach us about the folk conception of knowledge, since such lessons seem to depend upon the situations depicted being epistemically equivalent.

Despite the prima facie plausibility of this line of thinking, Nikolaus Dalbauer and Andreas Hergovich (2013) found that the goodness or badness of an outcome affects the
probability judgments that participants make about those outcomes and that it does so independently of whatever background beliefs participants might hold about how easy it is to harm or help the environment. Furthermore, when Beebe and Jensen (2012) modified the chairman and the environment vignette so that participants were told there was either a slight chance that the environment would be harmed or a very strong chance that it would be helped, participants were still more inclined to attribute knowledge in the harm condition than in the help condition. Together, these findings suggest that unbiased assessments of the relative probabilities of the good and bad outcomes are not what lie behind the ESEE.

A common explanation of the Knobe effect is that individuals’ motivation to blame the chairman for harming the environment has a distorting effect upon individuals’ intentionality attributions. Thomas Nadelhoffer (2006) and Mark Alicke (2008), for example, suggest that unconscious, spontaneous processes associated with blame attribution lead participants to exaggerate the chairman’s causal control over the negative environmental outcome and to lower the evidential standards for blaming him, which makes it easier to attribute intentionality to him. In other words, blame attribution processes distort intentionality attribution processes. Fred Adams and Annie Steadman (2004a, 2004b) also suggest that blame plays a central role in generating the Knobe effect, noting that an assertion of ‘You did that intentionally’ often serves the function of implicating that someone is to blame, even if the semantic content of ‘intentionally’ does not include blame as a component. In a similar fashion, it seems that ‘The chairman knew the environment would be harmed’ can implicate that he is to blame for his action. If denying that he knew this would happen generates the implicature that the chairman is not to blame for his action, this might increase knowledge attributions in the harm condition. Even if someone was not consciously engaged in implicating blame, unconscious blame
processes might nonetheless distort individuals’ judgments about who knows what, thereby producing the ESEE. Although this line of thinking has some initial plausibility, significant empirical challenges to it have been raised in the literature.

For example, Jonathan Schaffer and Knobe (2012, n. 10) report that they asked participants about the knowledge of a neutral, third-party observer that they added to Knobe’s original environment vignette, viz., “an environmentalist who knew that scientists were predicting helpful or harmful effects and then learned about the chairman’s decision to go ahead with the program.” Participants were more inclined to say the environmentalist knew that the environment would be harmed than they were to say that he knew it would be helped. Buckwalter (2014b) and Beebe (forthcoming) obtained similar findings. Blame ratings obtained by Knobe and Mendlow (2004), Mark Phelan and Hagop Sarkissian (2008), and Beebe (forthcoming) indicate that participants attributed knowledge to the relevant protagonists without always blaming them for their actions. Furthermore, Beebe (forthcoming) observed the ESEE when the believers in question were praiseworthy characters who nobly undertook significant personal costs. These findings do not comport well with the blame-based hypothesis under consideration. If an overactive motivation to blame agents leads individuals to attribute more knowledge to them than is warranted, we should not expect individuals to attribute the same level of knowledge to blameless observers or agents or to rate putatively blameworthy agents as blameless.

One hypothesis that purports to explain a wider range of data than the blame-based account is the ‘belief heuristic’ approach articulated by Mark Alfano, Beebe, and Brian Robinson (2012). They argue that agents whose actions violate normative expectations are more likely to
engage in deeper levels of reflection about the consequences of their actions than agents who do not violate such expectations. They write:

[T]rue beliefs to the effect that one is violating a norm are typically more valuable than true beliefs to the effect that one is conforming to a norm. One may be sanctioned for violating a norm, so forming a true belief about whether one has violated a norm (hence potentiating such a sanction) is valuable, regardless of whether one endorses the norm. The chairman in the HELP condition, for example, does not need to say to himself, “Wait! I need to stop and think carefully about whether helping the environment is something that I should be doing.” In the HARM condition, however, an inner monologue like this might well be appropriate. (Alfano, Beebe, and Robinson 2012, 269)

Increased reflection about consequences—even by praiseworthy agents—makes it more likely that agents who violate normative expectations will form stronger beliefs and come to have knowledge about the consequences of their actions and that attributors will reasonably interpret them as having done so. However, it remains to be seen how well the belief heuristic approach will fare as additional data comes in.

One way that the robustness of the ESEE has shown itself is that the badness of actions has been observed to override other considerations that upon reflection should militate against the attribution of knowledge. For example, Beebe and Shea (2013), Buckwalter (2014b), and Turri (2014) all found that participants were more willing to attribute knowledge in Gettier cases that included some kind of wrongdoing than in Gettier cases that did not. Beebe and Shea took the chairman and the environment case and ‘Gettiered’ it by giving the chairman a justified true belief but prevented the justification and truth from being related in the expected fashion. They also adapted well-known thought experiments from the Gettier problem literature and introduced
elements of wrongdoing. Across several sets of cases, Gettiered protagonists who performed bad actions were more likely to be seen as having knowledge than Gettiered protagonists in morally neutral situations. Buckwalter (2014b) and Turri (2014) obtained similar results.

However, Turri (2014) not only observed the ESEE in Gettier cases. He also observed it cases that involved justified false beliefs or no beliefs at all. In the latter cases, participants were still more inclined to attribute knowledge to an agent if the agent performed an action that was bad than if the agent’s action was morally good or neutral. The “virtually unlimited” scope of the ESEE leads Turri (2014, 107) to the following conclusion:

I think that it would make a conceptual-competence explanation of the observed effect much less likely. It would seem more likely that participants are incompetently applying their concept of knowledge, or competently but falsely applying it for pragmatic reasons, or competently applying some other concept in response to a “knowledge” question. Over thousands of years and across many different cultures, careful reflection has led people repeatedly to the view that knowledge requires truth and belief (or something very similar).

These concerns about the ESEE cause Turri to question the reliability of asking participants to indicate the extent to which they agreed or disagreed with a knowledge attribution on a five- or seven-point Likert scale ranging from ‘Strongly disagree’ to ‘Strongly agree.’ In an effort to see whether “an appropriately circumscribed” ESEE could be observed in “an overall pattern of knowledge attribution that broadly agrees with mainstream theorizing about knowledge,” Turri (2014, 117ff.) explored an alternative method. Using the same vignettes from previous experiments, Turri first asked participants to make a forced-choice between the options ‘really knows’ and ‘only thinks he knows’ before asking them to rate on a scale from 1 (‘not at all
confident’) to 10 (‘completely confident’) how confident they were in their answers to the forced-choice question. Scoring each ‘really knows’ answer as a +1 and each ‘only thinks he knows’ as a -1, and multiplying these values by participants’ confidence scores, Turri observed the ESEE in a pair of basic Knobe effect cases and a pair of justified false belief cases. He did not, however, observe the ESEE in a pair of Gettier cases—in contrast to his previous study. In addition, knowledge attributions in the false belief cases were a good bit lower than before. Because these results are more in line with traditional philosophical theorizing about knowledge, Turri (2014, 121) concludes not only that the alternative method he employed was superior to the previous method; he also concludes that the ESEE is “arguably eligible to inform substantive theorizing about knowledge or our concept thereof.”

If (i) one set of results in experimental philosophy is obtained using one method, (ii) a contrasting set of results is obtained using another method, and (iii) only one set of results contrasts sharply with a broad consensus among philosophers, I can understand why many philosophers would conclude that the more surprising results should be given less credence. However, experimental philosophy has been in the business of challenging and overturning established philosophical opinion. Turri (forthcoming a; forthcoming b) himself has argued—against overwhelming consensuses to the contrary—that the epistemic closure principle is not in fact a core part of folk epistemology and that skeptical challenges to ordinary knowledge claims are not based upon something that is terribly interesting from a philosophical perspective. Thus, the fact that Turri’s second set of results is more consistent with traditional philosophical theorizing may tell us very little about its trustworthiness.

Furthermore, there are a few features of Turri’s forced-choice question that raise some concerns. First, as Simon Cullen (2010) has shown, prompt questions that use ‘really knows’ can
lead participants to answer differently than questions that simply use ‘knows.’ Secondly and perhaps relatedly, the modifier ‘really’ may induce participants to use higher epistemic standards than they would otherwise, potentially distorting ordinary knowledge attributions. Thirdly, beliefs or opinions come in degrees of strength. A Likert scale matches this range of strength by presenting participants with an ordered range of answer choices. Instead of matching this feature of participants’ opinions, forced-choice questions require participants to translate a graded item into something categorical. Turri, of course, follows up his forced-choice question with a graded one about confidence. But this seems rather like translating an analog signal to a digital one, and then translating the digital signal back into analog, when there was never any need for the translation to occur in the first place. It seems preferable to let graded phenomena be represented by graded scales.

Nonetheless, Turri raises an important issue about the ESEE. It is often assumed that an unexpected effect must be viewed either as a reflection of folk conceptual competence or as a bias or distortion. Turri’s suggestion is that an effect like the ESEE might be a genuine reflection of competence in some situations but a performance error in others. This seems like a fruitful approach to take with respect to a wide variety of empirical findings.

The effects that non-epistemic evaluative judgments have on knowledge attributions raise important questions about the philosophical analysis of knowledge. Almost every epistemologist of the last fifty years has endorsed the idea that, since the target of standard analyses of knowledge is the ordinary person’s conception of knowledge, such analyses should be answerable to data about ‘what the ordinary person would say’ in response to various epistemological thought experiments. Indeed, most of the major movements and innovations in epistemology during the last fifty years have relied heavily upon intuitive responses to key
thought experiments. The ESEE, when viewed against the backdrop of this widely shared methodological commitment, suggests that the traditional JTB conception of knowledge may need to be revised further to accommodate non-epistemic evaluative considerations.

Proponents of pragmatic encroachment have already begun to challenge the traditional epistemological view that someone’s knowledge of a proposition is independent of whatever actions that person may undertake in light of believing that proposition. For example, Jeremy Fantl and Matt McGrath (2007, 559) defend the following knowledge-action principle:

\[(KA) \text{ } S \text{ knows that } p \text{ only if } S \text{ is rational to act as if } p.\]

John Hawthorne and Jason Stanley (2008) offer a similar principle:

\[(AKP) \text{ Treat the proposition that } p \text{ as a reason for acting only if you know that } p.\]

These principles are based upon the idea that non-epistemic facts about how much is at stake for a subject concerning the truth of the belief can affect whether the subject knows. The ESEE suggests a further respect in which knowledge may be connected to action, viz., that whether a subject’s belief counts as knowledge may be in part determined by other evaluative features of actions the subject performs in light of that belief. Adopting such a view would represent a significant departure from traditional philosophical theorizing about knowledge, but the method used to arrive at this suggestion is much less revisionary: ordinary individuals have been found to exhibit a robust tendency to take evaluative assessments of actions into account when making judgments about knowledge. These findings suggest that epistemological judgments may be more intertwined with other kinds of evaluative judgments than epistemologists have appreciated.
References


Buckwalter, Wesley, and Jonathan Schaffer. forthcoming. “Knowledge, Stakes and Mistakes.” *Noûs*.


