

Social Functions of Knowledge Attributions*

James R. Beebe (University at Buffalo)

Forthcoming in Mikkel Gerken and Jessica Brown (eds.), *Knowledge Ascriptions* (Oxford University Press)

Word count (not including notes and references): 8,677

Drawing upon work in evolutionary game theory and experimental philosophy, I argue that one of the roles the concept of knowledge plays in our social cognitive ecology is that of enabling us to make adaptively important distinctions between different kinds of blameworthy and blameless behaviors. In particular, I argue that knowledge enables us to distinguish which agents are most worthy of blame for inflicting harms, violating social norms, or cheating in situations of social exchange.

Keywords: knowledge, blame, genealogy, evolutionary game theory, experimental philosophy

Edward Craig (1986, 1990, 2007) has famously argued that new epistemological insights can be gained if philosophers begin to think about the point or purpose of having a concept of knowledge and what conceptual need its possession and use might satisfy. Directing epistemologists to imagine an epistemic ‘state of nature’ in which inhabitants are otherwise like

* I would like to thank Jessica Brown, Mikkel Gerken, and audience members at the 2010 Episteme Conference at University of Edinburgh and the 2010 Society for Skeptical Studies meeting at the Eastern division meeting of the American Philosophical Association for helpful comments and feedback on previous versions of this article.

us in their native cognitive endowments but lack our concept of knowledge, Craig (1990, p. 11) writes:

Human beings need true beliefs about their environment, beliefs that can serve to guide their actions to a successful outcome. That being so, they need sources of information that will lead them to believe truths. They have ‘on-board’ sources, eyes and ears, powers of reasoning, which give them a primary stock of beliefs. It will be highly advantageous to them if they can also tap the primary stocks of their fellows—the tiger that Fred can see and I can’t may be after me and not Fred—that is to say, if they act as informants for each other. On any issue, some informants will be better than others, more likely to supply a true belief.... So any community may be presumed to have an interest in evaluating sources of information; and in connection with that interest certain concepts will be in use. The hypothesis I wish to try out is that the concept of knowledge is one of them. To put it briefly and roughly, the concept of knowledge is used to flag approved sources of information.

Craig suggests that our modern-day concept of knowledge can be seen as developing from a more primitive concept (dubbed ‘protoknowledge’ by Kusch, 2009) that satisfied this need to indicate reliable sources of testimony.

When Craig (1990, p. 4) first put forward his genealogical account, he did not rule out the possibility that the concept of knowledge might play roles in our social cognitive ecology other than the one he describes:

Any society that has a well-developed language, sufficiently well developed for us to be able to say that it exercises a concept even approximately identifiable with our concept of knowledge, consists of creatures that have reached a considerable degree of mental

complexity. Any number of different sorts of need may, for all we know to the contrary, follow in the wake of this complexity; so there is no a priori reason to think that we are tied by methodological principles to considering only needs of the very basic kind that I have actually tried to restrict myself to.

Like a number of other scholars (e.g., Williams, 2002; Kusch, 2009, 2011; Kappel, 2010; Gelfert, 2011), I believe that a consideration of other aspects of our knowledge-generating and knowledge-consuming practices can reveal additional needs that the concept of knowledge might satisfy.

How does one ascertain what roles the concept of knowledge plays or what conceptual needs it satisfies? One of the characterizations Craig (1990, pp. 2, 3) provides of his ‘state of nature’ approach is the following:

We take some prima facie plausible hypothesis about what the concept of knowledge does for us, what its role in our life might be, and then ask what a concept having that role would be like, what conditions would govern its application.... And if [there is such a thing as the point of the concept of knowledge], one way to find out must be to form some hypothesis about it, try to work out how a concept custom-designed for that role would look, and then see to what extent it matches our everyday practice with the concept of knowledge as actually found.

According to this characterization of Craig’s methodology, the search for empirical evidence concerning the role the concept of knowledge plays enters at a relatively late stage in the process of genealogical reflection. From the armchair, so to speak, we formulate “some prima facie plausible hypothesis about what the concept of knowledge does for us” and only later do we “look, and then see to what extent it matches our everyday practice.” Despite the fact that Craig

thinks empirical evidence is required for this kind of project, little if any has actually been mustered in support of his genealogical account of knowledge. Such evidence as there is consists primarily of appeals to philosophical intuition. However, as Martin Kusch (2011) reminds us, because ecology is a science, “social cognitive ecology... should, wherever possible, reckon with the results of the natural and social sciences.”

An additional feature of Craig’s approach that keeps it from having a solid empirical grounding is that both the state of nature story he tells and the account he offers of how our concept of knowledge developed from an earlier concept of protoknowledge are purely hypothetical. Like many other state of nature stories, Craig’s is not hypothesized to represent any actual community in human history. As Bernard Williams (2002, p. 27) puts it, the state of nature is not the Pleistocene. Recently, however, several scholars (e.g., Williams, 2002; Kusch, 2009, 2011; Gelfert, 2011) have expressed dissatisfaction with the fact that hypothetical genealogies like Craig’s cannot tell us how and why the concept of knowledge actually emerged during the course of human evolution and have issued calls for the exploration of the development of the concept during the course of real human history.¹

In the present article I articulate and defend a view about one function the concept of knowledge performs in our social cognitive ecology that has a solid empirical grounding in recent findings from evolutionary game theory. On the view I defend, the concept of knowledge satisfies an important need that arises in the regulation of social interactions that can be summarized as follows:

Human beings need to engage in acts of social exchange—i.e., acts of cooperation for mutual benefit. That being so, humans need strategic information about other

¹ Readers might wonder how I can criticize Craig for certain purely hypothetical features of his work, while at the same time acknowledging that he recognizes a need for empirical support. For an insightful analysis of different strands of thought in Craig’s work that sometimes appear to be in tension, cf. Kappel (2010).

interactants. It will be highly advantageous if humans can distinguish not only the behavior of those who cooperate from those who do not but also the behavior of those whose failure to cooperate is intentional from those whose failure is unwitting. Being able to mark the latter distinction can prevent one from losing exchange partners who fail to reciprocate due to forces beyond their control—the goods that Fred promised me in return for my services may have been accidentally lost in a storm rather than willfully withheld—that is to say, if they punish or blame only those interactants who knowingly cheat. So any community may be presumed to have an interest in strategic information about the mental states of other interactants, and in connection with that interest certain concepts will be in use. The hypothesis I wish to try out is that the concept of knowledge is one of them. To put it briefly and roughly, the concept of knowledge is used to distinguish those agents who are most blameworthy for harm-inflicting, norm-violating or otherwise uncooperative behaviors.

This hypothesis is intended to supplement rather than to replace Craig's proposal.²

Instead of attempting to investigate the roles that the concept of knowledge plays from a purely hypothetical point of view and only later seeking empirical confirmation for my conjectures, I begin in section 1 by surveying recent findings from evolutionary game theory concerning the essential role that cooperative interaction has played in the course of human evolution and show that certain selection pressures seem to have favored the development of a concept of knowledge. In section 2 I chart several important conceptual connections between knowledge and blame and note that, other things being equal, the agents who are most

² The present hypothesis is also intended to be compatible with other proposals concerning the function of the concept of knowledge, including Kappel's (2010) and Kelp's (2011) claims that knowledge attributions are 'enquiry-stoppers,' although some delicate maneuvering may be required to specify how these different roles can be consistently reflected in the semantics and the pragmatics of knowledge attributions. I do not assume that every role the concept of knowledge plays will figure in some way in its truth conditions.

blameworthy for inflicting harm, violating social norms or making illicit assertions are those who know full well what they are doing, yet choose to do it anyway. In section 3 I examine a set of recent findings from experimental philosophy that concern ordinary attributions of knowledge and suggest that our apparent tendency to overattribute knowledge to blameworthy agents provides further experimental support for my hypothesis. I conclude there is strong reason to think that one of the functions the concept of knowledge performs is enabling us to draw adaptively important distinctions between different kinds of blame that we have a practical interest in attributing to agents who violate our normative expectations.

1. Knowledge and Social Exchange

Evolutionary game theory investigates strategies for interaction between individuals and assesses them on the basis of how they affect the Darwinian fitness of individuals pursuing them.³ Understanding strategies as specifications of what individuals will do in any situation in which they may find themselves and assuming that successful strategies will appear more often in the future, evolutionary game theorists formulate hypotheses about when individuals should cooperate, when they should be uncooperative, and the conditions under which different kinds of cooperation can arise and achieve stability within various populations (Maynard Smith, 1982; Axelrod, 1984). Some of the most basic lessons to emerge from evolutionary game theory include the following:

(1.1) It often pays to cheat (Trivers, 1971).

(1.2) Cheating or acting selfishly all of the time is generally less advantageous than cooperating with others at least some of the time (Axelrod, 1984).

³ Cf. Maynard Smith (1982) for an influential introduction to this field.

- (1.3) In many cases where the decision matrix for a single interaction favors uncooperative behavior, adding the assumption that interactants will encounter each other a large number of times in the future often results in cooperation becoming favored instead (Axelrod, 1984).
- (1.4) Humans are prone to cooperate, even with strangers (Richerson, Boyd and Henrich, 2003).
- (1.5) Displaying an initial willingness to cooperate is generally advantageous in establishing new, beneficial relationships (Axelrod, 1984).
- (1.6) In ongoing interactions with other individuals it is generally advantageous not to be the first one to defect or cheat (Axelrod and Hamilton, 1981; Axelrod, 1984).
- (1.7) Cooperating with everyone exposes one to exploitation by others (Axelrod, 1984).
- (1.8) Punishment of cheaters or defectors serves to make cheating unprofitable and to discourage future cheating (Axelrod, 1984; Clutton-Brock and Parker, 1995; Sober and Wilson, 1998).
- (1.9) People are willing to punish defectors, even at significant cost to themselves (Fehr and Gächter, 2002).
- (1.10) Punishing those who do not punish noncooperators or who do not punish nonpunishers can also be advantageous (Axelrod, 1986; Boyd and Richerson, 1992).
- (1.11) A single defection can sometimes lead to a chain of recriminations and counterrecriminations that results in suboptimal outcomes for all parties (Axelrod, 1984).

(1.12) Forgiving interactants who have previously defected or cheated but show a renewed willingness to cooperate is generally more advantageous than a policy of never forgiving (Axelrod, 1984).

(1.13) Natural selection favors psychological systems that can regulate both individuals' cooperative and uncooperative tendencies and their responses to these tendencies in others (Trivers, 1971).

(1.14) Natural selection favors the development of both subtler forms of cheating and more acute cheater-detection capacities (Trivers, 1971).

Many mutually beneficial behaviors (e.g., cleaning symbioses in fish and shrimp) do not require capacities for strategic reasoning (Trivers, 1971).⁴ However, more advanced strategies for interaction require increasingly sophisticated cognitive capacities. For example, if organisms are to move beyond strategies that treat all interactants the same, they need to be able to recognize particular individuals and to remember past interactions with them. Implementing some strategies (e.g., ones that advise doing whatever the other player does on each preceding move) only requires that organisms be able to remember a single previous interaction, while strategies that take into account wider ranges of prior interactions require greater memorial and computational resources.

It is widely agreed that social exchange—i.e., cooperation for mutual benefit—has been a crucially important feature in hominid evolution.⁵ Social exchange comes in a variety of different forms: barter, trade, gift giving, exchanging favors between friends, forming coalitions for mutual protection, alerting others to the presence of a food source, granting sexual access,

⁴ Axelrod (1984, p. 22) observes, "Cooperation in biological systems can occur... even when they are unable to appreciate the consequences of their own behavior."

⁵ Cf. Trivers (1971), Maynard Smith and Price (1973), Axelrod and Hamilton (1981), Maynard Smith (1982), Axelrod (1984), Cosmides (1989), Cosmides and Tooby (1989, 1997, 2005), Boyd and Richerson (1992), Sober and Wilson (1998), and Richerson, Boyd and Henrich (2003).

reciprocal grooming, and sharing information.⁶ In any situation of social exchange where repayment for benefits obtained is delayed, there will be opportunities for free-riding or cheating—i.e., for individuals to take benefits without reciprocating or otherwise fulfilling their end of the deal. Leda Cosmides, H. Clark Barrett and John Tooby (2010, p. 9008) write:

Using evolutionary game theory, it has been shown that adaptations for social exchange can be favored and stably maintained by natural selection, but *only if they include design features that enable them to detect cheaters*, and cause them to channel future exchanges to reciprocators and away from cheaters.⁷

Indiscriminate cooperation under conditions that allow cheating is an unstable strategy that will be selected against in all biologically plausible conditions. It is widely agreed that hominids experienced significant selection pressure toward developing minds that could detect cheaters. Cosmides (1989, p. 196) writes, “An individual who engaged in exchange, but who lacked the ability to detect cheaters, would experience fitness costs with no compensating benefits, and would be selected out.”

Most successful strategies advise punishing individuals who fail to cooperate. However, cognitively sophisticated agents who are capable of drawing distinctions between different kinds of failures to cooperate can gain an advantage over those who cannot draw such distinctions. Barrett (1999, p. 3) writes:

[I]t would be unfortunate to lose an exchange partner who failed to reciprocate because it was beyond his control (e.g. due to illness, bad luck in hunting, etc.), but who was operating in good faith and would reciprocate if possible. For this reason, we expect

⁶ Social exchange has been found in only a handful of species, including humans, chimpanzees, baboons, lions, and vampire bats, and the cognitive mechanisms underwriting the capacity for exchange develop in humans by ages 3-4 and appear crossculturally (Cosmides and Tooby, 2005, p. 1295).

⁷ Cosmides and Tooby (1989, pp. 56-57) write, “A cooperative strategy can invade a population of noncooperators if, and only if, it cooperates with other cooperators and excludes (or retaliates against) cheaters.”

mechanisms for cheater detection to be sensitive to more than just the fact that a benefit was taken and a cost not paid.

Thus, a strategy that denies future cooperation to agents who are blameworthy for prior failures of cooperation but remains open to cooperation with agents whose failures are due to factors beyond their control will outperform simpler strategies that refuse cooperation to all who have failed to cooperate. This more complex strategy, of course, requires agents to be able to make distinctions of blame and to reason about various psychological factors that determine blame.

Cosmides, Barrett and Tooby (2010, p. 9009) note that both intentional and accidental violations of social contracts involve one party receiving a benefit without paying for it, even though an intentional cheater represents a greater menace than an accidental one. Cheater detection systems, then, need to do more than simply detect compliance or noncompliance with a social contract—i.e., who receives which benefit and pays which cost. They need to make determinations of *mens rea*. That is, they need to determine whether acts of noncompliance are caused or underwritten by a guilty mind. Cosmides, Barrett and Tooby (2010, p. 9008) thus define a cheater not only as an agent who takes the benefit offered in a social exchange and fails to meet the provider's requirement (i.e., fails to pay for the benefit) but also as one who does so intentionally rather than by mistake or accident. Thinking about cheating, therefore, requires modeling the cost/benefit structure of a social exchange from more than one perspective (Cosmides and Tooby, 1989, p. 85; Cosmides and Tooby, 2005, p. 1297).⁸ Gerd Gigerenzer and Klaus Hug (1992, p. 132) write, "The notion of cheating implies that (at least) two *parties* with two different *perspectives* exist. Being cheated is relative to the perspective of one party."

⁸ Cosmides and Tooby (2005, p. 1297) write, "The definition of cheating is also perspective dependent, because the item or action that one party views as a benefit is viewed as a requirement by the other party. The system needs to be able to compute a cost-benefit representation from the perspective of each participant, and define cheating with respect to that perspective-relative representation."

Gigerenzer and Hug (1992), Cosmides and Tooby (1997), Fiddick (1998), Barrett (1999), and Cosmides, Barrett and Tooby (2010) all report experimental results in which the activation of cheater-detection algorithms crucially depend upon subjects being cued into the perspective of the party that can be cheated. Cosmides and Tooby (1997) and Fiddick (1998) found that when subjects were cued to take the perspective of looking for intentional cheaters, they were much better at identifying the “benefit taken” and “cost not paid” cards on a Wason selection task than when they were cued to the perspective of looking for “innocent mistakes” or unintentional violations of a social contract rule. Barrett (1999) found that an important factor in the activation of cheater detection was the extent to which subjects believed that potential cheaters had knowledge about possible ways to cheat. Full blown cheating, then, seems to require knowing what one is doing, and thus grasping something like the concept of knowledge appears necessary for forming the concept of a cheater (Malle and Knobe, 1997; Cosmides, Barrett and Tooby, 2010).

According to Klemens Kappel (2010), the heart of any Craigian ‘practical explication’ of knowledge consists in the advancement of the following two kinds of claims:

(2.1) Given a set of facts F [e.g., “about our physical environment, our biological set-up, our social organization”], and a set of aims or interests I , we have a certain need N .

(2.2) The concept of knowledge is what actually fulfils the need N .⁹

The central thesis of this paper is that, in addition to whatever need we may have for being able to flag approved sources of information, we also have the following need and the concept of knowledge plays an important role in fulfilling it:

⁹ In claiming that these two claims constitute the heart of a practical explication of knowledge, Kappel is in effect claiming that hypothetical genealogical stories are unnecessary.

(2.3) Given the importance of social exchange to human survival and flourishing and our interest in developing effective strategies for cooperative interaction, we have a need to distinguish those who are (and are not) most blameworthy for uncooperative behaviors.

My claim is that part of the point of having a concept of knowledge is that it enables us to fulfill this need by contributing to an understanding of what a cheater is.

I follow Kappel (2010, pp. 79-80) in understanding the nature of the relevant conceptual need in the following way:

The word 'need' should not be understood in the strong sense according to which S has a need for X in order to do F only if S cannot do F without X . In this sense, humans have a need for water, for example, since without water humans die. When I here speak of a conceptual need I have a less dramatic sense of the word 'need' in mind; my suggestion is only that, other things being equal, we are better off if we can meet this conceptual need in some way. We are better off meeting the need, though we might also get along without doing so.

I should also note that my claim is not that knowledge is the only possible human concept that could have satisfied the need in question. Using the example of a practical explication of the point of having cars, Kappel (2010, p. 72) writes:

[T]his explanation is not meant to explain why we have an extensive traffic system based on privately owned vehicles *rather than*, say, a very elaborate system of public transport. Like many other explanations, the practical explication of cars depends on contrast classes. Roughly, saying that we have cars in order to meet a need for transport is meant to explain why we have cars, rather than, say, no means of transport apart from our

natural ability to walk and run. It is not meant to explain why we have cars rather than some other equally efficient system of transport.

Accordingly, while it might be possible for an array of concepts other than knowledge to satisfy the need in (2.3), my claim is that the concept of knowledge has been the one that has actually satisfied it.¹⁰

Finally, (2.3) should not be interpreted as implying that every deployment of the concept of knowledge is concerned in some way with blame. If the concept of knowledge had only one function or fulfilled only one need, then perhaps proper deployment of the concept might require something like this. However, if knowledge fulfills multiple needs, there will likely be circumstances in which its use succeeds in fulfilling one kind of need (e.g., flagging approved sources of information) without failing to fulfill others (e.g., making distinctions of blame or signaling that a particular line of enquiry has been pursued far enough¹¹).

In contrast to Craig's proposal, which appeals to a merely hypothetical genealogy and for which little empirical evidence is available, I contend that the current proposal, which is rooted in evolutionary game theory, rests on a more solid empirical footing. In the sections that follow I elaborate the foregoing hypothesis by exhibiting some important conceptual connections that exist between attributions of knowledge and attributions of blame (section 2) and surveying recent experimental findings that reveal a robust tendency to overattribute knowledge to blameworthy agents (section 3).

¹⁰ Thanks to Jessica Brown for bringing this point to my attention.

¹¹ Cf. Kappel (2010).

2. Knowledge and Blame: Conceptual Connections

To appreciate the role that knowledge plays in determining the degree of blame an agent should receive for bringing about a harm, let ‘*W*’ denote a warrant-giving ground of belief—i.e., a ground that suffices to make any true belief count as knowledge. And let ‘*J*’ denote a ground of belief that confers positive epistemic status upon a belief but nevertheless fails to make the belief, if true, count as knowledge. Now consider the following possible psychological states of an agent, *S*, whose action, *A*, has resulted in a harm, *H*:

(3.1) *S* did not know that *A* would result in *H* because (i) *S* did not believe *A* would result in *H* and (ii) *S* possessed no *J* or *W* for believing *A* would result in *H*.

(3.2) *S* did not know that *A* would result in *H* because (i) *S* did not believe *A* would result in *H*, although (ii) *S* possessed some *J* (but no *W*) for believing *A* would result in *H*.

(3.3) *S* did not know that *A* would result in *H* because (i) *S* did not believe *A* would result in *H*, although (ii) *S* possessed some *W* for believing *A* would result in *H*.

(3.4) *S* did not know that *A* would result in *H* because, although (i) *S* believed *A* would result in *H*, (ii) *S* possessed no *J* or *W* for believing *A* would result in *H*.

(3.5) *S* did not know that *A* would result in *H* because, although (i) *S* believed *A* would result in *H*, (ii) *S* possessed some *J* (but no *W*) for believing *A* would result in *H*.

(3.6) *S* knew that *A* would result in *H*.

Surely it is the case that, other things being equal, the psychological state described in (3.6) is the most blameworthy one for *S* to be in and (3.1) the least blameworthy, with the others falling somewhere in between, roughly in the order in which they are presented.¹² The most

¹² Although it seems clear to me that *S* would deserve more blame in (3.6) than in (3.5), a case might be made that if the only difference is that *S* is in a Gettier-style situation in (3.5) but not in (3.6), *S* should be equally

blameworthy wrongdoers are those who ‘know full well’ that they are doing wrong, yet choose to do it anyway. In American English ‘knowing full well’ is invariably used to express the highest degree of blame possible for a given action. In attributing the possession of complete knowledge—not mere belief, merely true belief or merely justified belief—it suggests that the agents had no epistemic excuse for doing what they did. In making distinctions of blame, it is important to discriminate between those who have informational access to certain facts and those who do not—and to the quality or kind of informational access they have and the deliberative use they make of that information.

In a similar fashion, knowledge also plays a role in determining the degree of blame an agent should receive for violating a social norm, *N*. Consider:

(4.1) *S* did not know that *A* would violate *N* because (i) *S* did not believe *A* would violate *N* and (ii) *S* possessed no *J* or *W* for believing *A* would violate *N*.

(4.2) *S* did not know that *A* would violate *N* because (i) *S* did not believe *A* would violate *N*, although (ii) *S* possessed some *J* (but no *W*) for believing *A* would violate *N*.

(4.3) *S* did not know that *A* would violate *N* because (i) *S* did not believe *A* would violate *N*, although (ii) *S* possessed some *W* for believing *A* would violate *N*.

(4.4) *S* did not know that *A* would violate *N* because, although (i) *S* believed *A* would violate *N*, (ii) *S* possessed no *J* or *W* for believing *A* would violate *N*.

(4.5) *S* did not know that *A* would violate *N* because, although (i) *S* believed *A* would violate *N*, (ii) *S* possessed some *J* (but no *W*) for believing *A* would violate *N*.

(4.6) *S* knew that *A* would violate *N*.

deserving of blame in both cases. I won’t take the time to contest this suggestion, since nothing essential to my case depends upon it, but cf. the discussion below on the fact that the concept of knowledge is the easiest epistemic concept for folk attributors to deploy. Thanks to Jessica Brown and Mikkel Gerken for bringing this point to my attention.

As in (3.1) through (3.6), the worst kind of norm violator is represented in (4.6), with the others being ordered roughly in terms of increasing levels of blame. Norm violators who knowingly and intentionally violate norms are subject to greater blame than those who unwittingly do so.

Although the philosophical literature on the knowledge norm of assertion has focused upon the kinds of assertion one would be least blameworthy in making, consider the following illicit assertions—i.e., cases in which *S* asserts that *p* but one of the following is true:

(5.1) *S* did not believe that *p*, and *p* was true.

(5.2) *S* did not believe that *p*, and *p* was false.

(5.3) *S* believed that not-*p*, *p* was true, and *S* possessed no *J* or *W* for believing that not-*p*.

(5.4) *S* believed that not-*p*, *p* was false, and *S* possessed no *J* or *W* for believing that not-*p*.

(5.5) *S* believed that not-*p*, *p* was true, and *S* possessed some *J* (but no *W*) for believing that not-*p*.

(5.6) *S* believed that not-*p*, *p* was false, and *S* possessed some *J* (but no *W*) for believing that not-*p*.

(5.7) *S* knew that *p* was false.

It seems that, other things being equal, *S* violates more norms, violates them more seriously, and hence is deserving of more blame if (5.7) is true than if any of the other descriptions are true. The violations grow more severe as *S*'s epistemic position with respect to not-*p* grows stronger. Regardless of whether knowledge is capable of distinguishing the most blameless kind of assertion, it does seem capable of distinguishing (one of) the most blameworthy kind(s).¹³

¹³ Although I will not take time to argue the point here, I take the connection between knowledge and blameworthiness to be more fundamental than whatever connection (if any) exists between knowledge and completely blameless assertions, actions and instances of practical reasoning. Craig (1990, p. 49) presages discussion of the knowledge norm of assertion with the following analogy of what it is like to have relied upon

In each of the above lists—i.e., (3.1) through (3.6), (4.1) through (4.6) and (5.1) through (5.7)—the clearest and easiest to grasp distinction is that between agents who have knowledge about key features of their actions and those who do not. Although epistemologists and others whose work requires them to think carefully about knowledge and evidence (e.g., scientists, trial lawyers and judges) may be able to draw a variety of subtle distinctions between an array of distinct epistemic statuses, we can expect folk epistemic assessments to be framed in terms of common, accessible concepts such as knowledge.¹⁴ Thus, a variety of agents can be expected to appreciate the following ways in which knowledge can affect the degree of blame an agent deserves:

(6.1) Other things being equal, if *S*'s action results in a harm, *S* is more deserving of blame if *S* knows the action will result in the harm than if *S* does not.

(6.2) Other things being equal, if *S*'s action violates a social norm, *S* is more deserving of blame if *S* knows the action will violate the norm than if *S* does not.

(6.3) Other things being equal, if *S* asserts that *p*, *S* is more deserving of blame if *S* knew that *p* was false than if *S* did not.

someone's assertion, only to discover that the asserter's belief was accidentally true: "It produces that retrospective feeling of having run a risk, of having done something that one would not have done had one just been a little better informed at the time, rather like finding that the person who has just driven you 50 miles down a busy motorway without incident hasn't passed the driving test." The idea that nothing short of non-accidentally true belief can fully warrant testimony has been taken up in the idea that knowledge is the constitutive norm of assertion—i.e., that one must assert that *p* only if one knows that *p* (cf. Williamson, 1996). Many defenders of the knowledge norm of assertion contend that it may simply be a special case of a more general norm concerning knowledge and action. Stanley (2005, p. 9), for example, maintains "it is immensely plausible to take knowledge to be constitutively connected to action, in the sense that *one should act only on what one knows*," and Hawthorne (2004, p. 30) has argued, "if the question whether *p* is practically relevant, it is acceptable to use the premise that *p* in one's deliberations if one knows it and (at least in very many cases) unacceptable to use the premise that *p* in one's practical reasoning if one doesn't know it. ." More recently, Hawthorne and Stanley (2008) have defended the following principle concerning knowledge and practical reasoning: "Treat the proposition that *p* as a reason for acting only if you know that *p*."

¹⁴ As Pinillos (in press) notes, 'knowledge' is not only one of the most commonly used mental state words in the English language, it is also one of the most commonly used mental state words among young children. For further support of these claims, cf. Shatz, Wellman and Silber (1983) and Oxford English Corpus: <http://www.oxforddictionaries.com/page/oecfactslanguage/the-oec-facts-about-the-language>.

Therefore, the concept of knowledge can be seen to play an important role in distinguishing agents who deserve the highest levels of blame from those who do not.

Granted, there are many other folk psychological concepts (e.g., intention, desire, etc.) that are also relevant to assessments of blame, and the same selection pressures that are hypothesized to have favored the development of the concept of knowledge will also likely have favored their development as well. I take this consequence to be a strength of my position, since it would be problematic if an account of the development of the concept of knowledge were unable to interact with accounts of how the rest of our folk psychology developed.¹⁵

3. Overattributing Knowledge to Blameworthy Agents

Recent investigations in experimental philosophy have found that people are strongly inclined to count as knowledge the beliefs of agents who are blameworthy for actions undertaken in light of those beliefs—even when the epistemic positions of those agents do not seem to warrant such attributions and even when the attributors would not otherwise be inclined to attribute knowledge to them. The hypothesis that part of the reason we have a concept of knowledge is to enable us to distinguish those agents who are most blameworthy for violating our normative expectations provides a ready explanation for these behaviors, whereas the hypothesis that the sole function of the concept of knowledge is to enable us to flag approved sources of information renders these data surprising and perhaps indicative of widespread bias or irrationality.

In contrast to the received view in epistemology that whether a true belief counts as knowledge is independent of whatever actions a believer might undertake in light of her belief, my collaborators and I have recently found that ordinary participants' assessments of actions can have a significant effect on the epistemic assessments they make of an agent's beliefs. For

¹⁵ Cf. Morton (2001) for defense of a similar view.

example, Beebe and Buckwalter (2010) and Beebe and Jensen (in press) discovered that experimental participants are more likely to say that an agent knew that her action would bring about a certain side-effect if the side-effect is bad than if it is good.¹⁶ Beebe and Buckwalter reported on a between-subjects experiment in which participants were given 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. We are sure that it will help us increase profits, and 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 *helped/harmed*.

Participants were asked to indicate their response to the question “Did the chairman know that the new program would *help/harm* the environment?” on a seven-point Likert scale, ranging from -3 (labeled as ‘the chairman didn’t know’) to 3 (labeled as ‘the chairman knew’). 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%), and the percentage of participants who chose responses 1, 2 or 3 in the harm case (90%) was significantly greater than the number of participants who chose 1, 2 or 3 in the help case (61%).¹⁷

Beebe and Jensen (in press) found the same pattern of results in cases that do not involve environmental harm or the violation of ethical norms. For example, participants who read about a movie studio executive who was thinking about implementing a new policy that would increase profits but would make the studio’s movies worse from an artistic perspective tended to strongly

¹⁶ Each of the studies below builds upon Knobe’s (2003a) pioneering work in experimental philosophy on folk attributions of intentionality. Similar results have been found for attributions of desire, belief, being in favor of, and deciding (cf. Alfano, Beebe and Robinson, in press).

¹⁷ Beebe and Buckwalter call this asymmetric pattern of responses ‘the epistemic side-effect effect.’

agree with the statement “The CEO knew that the new policy would make the movies worse from an artistic standpoint.” However, participants were significantly less inclined to think that an otherwise identical movie studio executive knew that his new policy would make movies that were better.¹⁸ Beebe and Jensen also found that participants were inclined to say that the head of a computer corporation knew that her corporate restructuring plan would decrease sales in New Jersey for the following quarter, while participants in a parallel condition did not think that she knew that her plan would increase sales in New Jersey.¹⁹

In each of these cases, an agent in a decision-making role receives testimony about the potential side-effect of a certain course of action. In both the help and the harm conditions it seems that the agent possesses epistemically equivalent justification for believing the side-effect will occur, yet there are systematic differences in how participants respond in the two conditions. These differences have led many researchers to suggest that folk psychological judgments may have functions other than yielding value-neutral reports of perceived psychological facts.

When the effects of good and bad side-effects on folk psychological judgments first came to light, Joshua Knobe (2003a, 2003b, 2004a, 2004b) maintained that there was something proper about these judgments being influenced by factors that most other philosophers would regard as irrelevant to the correct application of the target concepts. Knobe argued that when value judgments of various kinds cause folk psychological attributions to go in unexpected directions—directions they would not have gone, were it not for those value judgments—we should not view this process as one of distortion. Instead, he suggested we should broaden our

¹⁸ As in the chairman and the environment case, nearly twice as many participants selected the strongest affirmation of the CEO’s knowledge (i.e., response ‘3’) in the ‘worse’ condition (48.7%) as in the ‘better’ condition (26.1%), and the overall distributions in the two conditions were significantly different from one another.

¹⁹ Nearly twice as many participants chose response ‘3’ in the harm condition (44.3%) as in the help condition (23.2%), and the overall distributions in the two conditions were significantly different from one another.

conception of the point and purpose of folk psychological judgments. For a time Knobe (2007) seemed to disavow this view, but has more recently reaffirmed his commitment to it:

One approach [to explaining the experimental data] would be to say that the relevant competencies are entirely non-moral but that some additional factor (conversational pragmatics, performance error, etc.) then interferes and allows people's moral judgments to affect their intuitions. Another approach would be to say that moral considerations truly do figure in workings of the competencies themselves. I argue that the data available now favor the second of these approaches over the first. (2010, p. 315)

Knobe (ibid.) sets himself against the first view, which he further characterizes as the view that "people truly are engaged in an effort to pursue something like a scientific investigation, but that they simply aren't doing a very good job of it" because, although "the competencies underlying people's judgments actually are purely scientific in nature,... there are then various additional factors that get in the way of people's ability to apply these competencies correctly."²⁰

Translating Knobe's reflections into the terms of the debate surrounding the point of having a concept of knowledge, we might recharacterize the view that "people truly are engaged in an effort to pursue something like a scientific investigation, but that they simply aren't doing a very good job of it" as the view that since the sole point of having a concept of knowledge is to enable us to mark approved sources of information, any use of the concept that does not perform

²⁰ Other statements by Knobe on the same subject include the following:

To some degree at least, it seems that these results should come as a surprise to those who think of people's concept of intentional action as a tool for predicting, controlling and explaining behavior. After all, it seems that the best way to accomplish these 'scientific' goals would be to ignore all the moral issues and focus entirely on a different sort of question (e.g., on questions about the agent's mental states). How then are we to make sense of the fact that moral considerations sometimes influence people's application of the concept of intentional action? (2006, p. 207)

Law courts commonly assume that judgments of purpose are purely factual judgments to be decided by juries. The same assumption is made in the literature on the child's theory of mind. The side-effect effect suggests, however, that such judgments may sometimes be partly factual and partly moral. To the extent that such judgments are moral, theory of mind is unlike a scientific theory, and its development is not reducible to discovering matters of fact. (Leslie, Knobe and Cohen, 2006, p. 426)

this function properly must be in error. The alternative view suggested by Knobe seems quite amenable to my own proposal, which states that since there is a multiplicity of needs the concept of knowledge fulfills, attributions of knowledge to agents who harm the environment, make movies worse or decrease sales in New Jersey (and denials of knowledge to agents who help the environment, make movies better or increase sales in New Jersey) might not constitute errors simply because they are not doing a good job of marking approved sources of information. They may well be doing a perfectly fine job of serving some other purpose instead. (It should be noted that knowledge attributions and denials can perform the hypothesized functions associated with blame attribution without actually being true. The linguistic norms governing established but in some sense alternative uses of folk psychological concepts might be encoded in the pragmatics of language use without figuring in the truth conditions for these statements.)

Further evidence that the folk are inclined to overattribute knowledge to blameworthy agents can be found in the work of several of Knobe's critics. Fred Adams and Annie Steadman (2004a, 2004b, 2007), for example, argue that asymmetric attributions of intentional action in side-effect cases do not indicate anything about the folk's core concept of intentional action. Instead, they suggest that due to pragmatic features of language use, saying 'You did that knowingly' or 'You did that on purpose' is a social way of assigning blame and discouraging actions of which one disapproves. Adams and Steadman hypothesize that people are inclined to call the chairman's action in the harm case above 'intentional' because denying it was performed intentionally would generate the unwanted implicature that he deserved no blame for harming the environment. They also suggest that subjects in the help condition are disinclined to say that the chairman performed his action intentionally because this might generate the unwanted implicature that he deserved praise for helping the environment. Adams and Steadman (2004a)

also predicted that giving participants the option to say that the chairman ‘knowingly’ harmed the environment would allow them to express disapproval of his action in the same way that saying he performed it intentionally would and that, since expressing disapproval is something they are likely to be motivated to do, participants would readily select this option. Beebe and Buckwalter (2010) found some confirmation for this prediction in a random selection of general newspaper articles archived on LEXIS-NEXIS, in which 62% of non-duplicate uses of ‘knowingly’ were associated with (typically criminal) wrongdoing, as in “the suspects knowingly used false Social Security cards” or “the leaders knowingly allowed a travesty.”

Further experimental evidence linking attributions of knowledge and blame comes from Shaun Nichols and Joseph Ulatowski (2007), who presented participants with the original chairman and the environment case described above and asked them whether the chairman helped or harmed the environment intentionally. Nichols and Ulatowski then asked participants to explain why they answered the question about intentional action as they did. Typical answers from participants who thought that the chairman intentionally harmed the environment include the following:

(7.1) He knew the consequences of his actions before he began.

(7.2) He knew that implementing the new program would hurt the environment.

(7.3) Because he knew he was going to hurt the environment, but chose to do it anyway.

(7.4) The chairman knew that this program would harm the environment.

Typical answers from participants who said that the chairman did not intentionally help the environment include:

(8.1) He didn’t care. It was an unintended consequence.

(8.2) Because his intention was to make money whether or not it will help the environment.

(8.3) He didn't INTEND on helping the environment, he INTENDED on making a profit.

(8.4) His motive was to get as much money as possible.

Thus, participants who thought that the chairman intentionally did something wrong and was blameworthy for doing so typically explained their answers in terms of the chairman's prior knowledge of the outcome, whereas those who did not think he did something intentionally or was blameworthy did not explain their answers in terms of knowledge.

In another study Thomas Nadelhoffer (2006) found that the blameworthiness of agents affected whether participants thought certain fatal outcomes had been knowingly brought about. In one story a thief who speeds off from a red light while a police officer holds on to the side of his car causes the car to swerve in a zigzag fashion in the hope of shaking the officer off. The thief knows full well that he is placing the officer in grave danger. His attempt to shake off the officer is successful, and the officer dies after being struck by another vehicle. In a second story a civilian speeds off from a red light after a thief brandishes a gun in his window. The thief holds on to the side of the car, and the driver again swerves in a zigzag fashion in the hope of escaping, knowing full well that doing so places the thief in grave danger. The attempt to shake off the thief is successful, and the thief is struck by oncoming traffic and dies. When asked "Did the thief knowingly bring about the officer's death?" 75% of participants who read the first story answered "Yes," while only 51% who read the second story answered "Yes" to "Did the driver knowingly bring about the thief's death?" Because the thief's actions in the first story is more blameworthy than the driver's actions in the second, it is plausible to think that blame is the factor that leads participants to be more inclined to attribute 'knowingly' to the first thief.

Some theorists (e.g., Alicke, 2008) argue that blame attributions have an especially strong tendency to influence other psychological attributions when participants are first told about a bad outcome and then asked to make a judgment about the agent's state of mind prior to performing the action that led to the bad result. Beebe (in press) obtained data supporting this contention by presenting participants with truncated versions of the environment, movies, and New Jersey vignettes described above that did not include information about the outcomes in question. The last two sentences of the original chairman and the environment vignette are "They started the new program. Sure enough, the environment was *helped/harmed*." These sentences and their analogues in the other cases were deleted, so that participants only knew which outcomes were predicted to occur, not whether they actually occurred. In each case the pattern of responses was significantly different from prior cases in which participants knew whether the predicted outcomes actually came to pass. Participants in the "no hindsight information" conditions were not significantly more inclined to attribute foreknowledge to the chairman and other protagonists in the harm conditions than participants in the help conditions. These results suggest that participants are basing their attributions of knowledge upon things other than the evidence available to the protagonists at the time of decision. If participants focused only upon the fact that seemingly reliable, epistemically equivalent testimony was given to the protagonists in the help and the harm conditions, knowing the ultimate outcome should not affect their epistemic assessments. Processes associated with blame attribution are plausible candidates for the intervening factors that make the relevant difference.

In another study Beebe (in press) followed a suggestion from Jonathan Schaffer and Knobe (in press) and placed a neutral, third-party observer in several of the side-effect cases. In the chairman and the environment vignette an environmentalist was added:

The CEO of a large company was thinking of starting a new program. An environmentalist closely monitoring the situation learned that scientists were reporting that the new program would *help/harm* the environment. The environmentalist then learned that the CEO had decided to start the new program. After the program was started, sure enough, the environment was *helped/harmed*.

Participants were then asked whether the environmentalist knew that the new program would *help/harm* the environment. Similar observers (e.g., journalists or other employees) were introduced into the movie studio and New Jersey cases described above. Because the neutral observers were clearly not agents that could properly be blamed if the environment, movies or sales were harmed, the prediction was that if perceived blameworthiness was actually playing a role in shaping participants' asymmetric knowledge attributions in side-effect cases, there should be a significant decrease in knowledge attributions in the new harm cases. Significant differences between participant responses in the original cases and in the modified neutral-observer cases were found. Regardless of whether outcome information (e.g., "Sure enough, the environment was helped/harmed") was included or not, participant responses did not differ significantly in the help and harm conditions. Like the central protagonists of the vignettes, the neutral observers had epistemically equivalent reasons to believe that the predicted outcomes would come about in both the help and the harm versions. However, unlike those protagonists, the observers did not perform blameworthy actions in the harm conditions. Thus, it is plausible to think that differences in blame processing could be responsible for the differences in participant responses.²¹

²¹ Beebe and Jensen (in press) also found that blame processing in some cases might have a greater overall effect on folk attributions than even explicit information about the likelihood that certain side-effects will come about. They presented participants with the same truncated vignettes pairs described above, except that each bad or harmful side-effect was to have a slight chance of occurring, while each good or helpful side-effect was said to have

The foregoing results establish the following predictive relation between attributions of knowledge and blame:

(9.1) The more blameworthy an agent is viewed as being for bringing about a given side-effect, the more likely attributors will be to think the agent knew the side-effect would occur.

It is important to appreciate the restricted scope of this claim, as well as its probabilistic and descriptive nature. (9.1) does not purport to apply to every situation where knowledge or blame are involved, nor does it guarantee that knowledge will be attributed to every agent who is blameworthy for bringing about some side-effect. Because agents will differ in the strength of their epistemic positions with respect to the propositions in question, there will be variations in how likely others will be to attribute knowledge to them. Individual differences in attributors' general willingness to attribute knowledge will likely have an influence as well.²² Furthermore, as a psychological claim, (9.1) is purely descriptive. It does not aim to specify conditions for the correct attribution of knowledge. It simply makes the claim that, other things being equal, ordinary subjects who attribute certain kinds of blame to an agent will be more likely to attribute certain kinds of knowledge to that agent. Beebe and Shea (in press) uncovered evidence that this claim holds even in cases where central epistemic features of an agent's situation seem to call for the withholding attributions of knowledge because the agent's situation has been "Gettierized." In other words, participants who would not otherwise be inclined to attribute knowledge to an

a very strong chance of occurring. Participants were then asked they thought central protagonists knew that the predicted outcomes would occur. The prediction was that, if participants are more strongly influenced by explicit information about objective probabilities, we should expect higher attributions of knowledge in each of the help conditions, but if participants are more influenced by the goodness or badness of the outcomes in question, we should find higher knowledge attributions in the harm conditions. In all four cases participants attributed knowledge to the chief protagonists in significantly greater numbers in the 'slight chance harm' conditions than in the 'very strong chance help' conditions.

²² Buckwalter and Stich (in press), for example, claim to have found evidence that women are in general more inclined than men to attribute knowledge to others.

agent become strongly inclined to do so if they take the agent to be blameworthy for bringing about certain kinds of outcomes.

A factor closely associated with blame that is commonly cited in explanations of side-effect results is the violation of salient social norms. Knobe (2007) briefly entertained—but no longer seems to endorse²³—the hypothesis that the asymmetries in folk psychological attributions in side-effect cases might be due to a nonconscious cognitive process dedicated to the detection of norm violations. Knobe hypothesized that the consideration of an action is always accompanied by an immediate, normative assessment of the action that takes into account only the most salient social norm, occurs below the level of conscious awareness, involves a very shallow level of processing, and does not involve any reflective considerations about whether the action is justified. The process is focused solely on detecting whether or not the action violates the norm in question. The nonconscious judgment then remains in memory and can affect other folk psychological judgments the subject might make. In particular, Knobe hypothesized that activation of the norm violation detection process increases the likelihood that folk psychological attributions of intentionality (and, by extension, knowledge) will be made.²⁴

Knobe's 'transgression-detection' model appears capable of explaining not only many of the results described above but also some more recent findings that are often taken to pose difficulties for blame-based models. For example, Knobe (2007) and Beebe and Jensen (2010) found the familiar asymmetric pattern of responses in intentionality and knowledge attributions when participants were given either the 'violate' or 'fulfill' version of the following vignette:

²³ Knobe discusses the 'transgression detection' model only in his 2007 paper and makes no mention of it in his more recent writings.

²⁴ Other researchers who have different explanatory models that are nonetheless centered around norm violations include Uttich and Lombrozo (2010) and Holton (2010).

In Nazi Germany, there was a law called the ‘racial identification law.’ The purpose of the law was to help identify people of certain races so that they could be rounded up and sent to concentration camps. Shortly after this law was passed, the CEO of a small corporation decided to make certain organizational changes. The vice-president of the corporation said: “By making those changes, you’ll definitely be increasing our profits. But you’ll also be *violating/fulfilling* the requirements of the racial identification law.” The CEO said: “[Look, I know that I’ll be *violating/fulfilling* the requirements of the law, but] I don’t care one bit about that. All I care about is making as much profit as I can. Let’s make those organizational changes!” As soon as the CEO gave this order, the corporation began making the organizational changes.

81% of participants in the ‘violate’ condition of Knobe’s study said that the CEO violated the requirements of the law intentionally, but only 30% in the ‘fulfill’ condition said that he fulfilled the requirements intentionally. Beebe and Jensen deleted the chairman’s self-attribution of knowledge (i.e., “Look, I know that I’ll be *violating/fulfilling* the requirements of the law”) and found that 48% of participants in the ‘violate’ condition affirmed in the strongest way possible on a seven-point scale that the CEO knew he would be violating the law, whereas only 23.5% of participants in the fulfill condition chose a similar response. The overall distribution of responses was also significantly different in the two conditions.

Violating the racial identification law in the present case is a good thing because fulfilling it would mean participating in a process that sends innocent people to concentration camps. However, because the issue is framed in terms of a legal violation, and legal violations are norm violations, Knobe’s prediction that the side-effect effect would be found was thus confirmed. These results are widely taken to pose a challenge for blame-based theories that predict increased

knowledge and intentionality attributions when agents are more blameworthy because the blameworthy CEO who fulfilled the unjust law was not seen as intentionally fulfilling the law or knowing that his actions would do so, whereas the blameless CEO who violated the law was. Although the all-things-considered blamelessness of the CEO in the Nazi case is generally taken to pose a serious (if not decisive) challenge to blame-based accounts of side-effect results, this conclusion is not in fact warranted by the available data. Knobe's norm violation model posits the existence of an automatic, unconscious cognitive process that is responsible for the detection of norm violations that is not under voluntary, executive control. However, the judgment that the CEO (and his actions) in the Nazi case are blameless is very likely the result of a conscious, reflective process, since it requires serious task interpretation and the recognition that an action that is normally bad (i.e., breaking the law) is balanced off by a greater good (i.e., not contributing to the deaths of innocent people). The fact that explicit measures of a conscious process fail to uncover any sign of blame attribution tells us nothing about whether the outputs of a distinct, nonconscious process involve blame-related considerations. Different (perhaps implicit) measures are required.²⁵

In any case, results from experiments using norm violation cases provide reason to think the following predictive relation obtains between knowledge attributions and norm violations:

(9.2) Attributors are more likely to think that agents who have violated salient social norms knew that their actions violated the norms than they are to think that agents who have fulfilled (or at least did not violate) norms knew that their actions did so.

²⁵ For further information on the contrast between implicit and explicit processes and their importance in recent psychological theorizing, cf. Evans and Frankish (2009).

Of course, since agents who violate norms are generally subject to some form of blame, the current connection between knowledge attributions and norm violations is not unrelated to the connection between knowledge and blame discussed above.

The vast majority of explanations offered to explain the data described in this section depict folk epistemic judgments as subject to the distorting influence of seemingly irrational biases to a worrisome degree.²⁶ On the view I am proposing, however, regardless of whether the robust tendency to overattribute knowledge is irrational (when judged solely from the perspective of marking approved sources of information), it should at least be unsurprising because part of the reason we have a concept of knowledge is that it enables us to distinguish which agents are most worthy of blame for inflicting harms, violating social norms, or otherwise acting in an uncooperative manner. When subjects are inclined to attribute blame, it is natural to expect them to be inclined to attribute underlying psychological states that are commonly associated with or determinative of blame.

4. Conclusion

I have articulated a pluralistic approach to the question of the point or purpose of having a concept of knowledge by granting that something like Craig's proposal is probably correct as far as it goes while at the same time arguing for a complementary alternative. Drawing upon work in evolutionary game theory, I have argued that key features of strategies for engaging in successful cooperative interaction reveal that an ability to deploy the concept of knowledge confers a selection advantage on those who possess it and that this in turn provides part of the explanation for why we have the concept. If we did not have the concept of knowledge, we

²⁶ Cf. Beebe and Buckwalter (2010) and Alfano, Beebe and Robinson (in press) for detailed discussions of the most prominent explanations of these data.

would not be able to draw practically important distinctions between different kinds of harm-doers, norm violators or uncooperators and the varying degrees of blame they deserve. By grounding my proposal in widely accepted views about mechanisms that have been important in human evolutionary history, I hope to encourage other social cognitive ecologists to pursue more empirically informed lines of inquiry as well.

References

- Adams, F. & Steadman, A., 2004a. Intentional action in ordinary language: core concept or pragmatic understanding?, *Analysis* 64, pp. 173-81.
- Adams, F. & Steadman, A., 2004b. Intentional action and moral considerations: still pragmatic, *Analysis* 64, pp. 268-76.
- Adams, F. & Steadman, A., 2007. Folk concepts, surveys, and intentional action. In: C. Lumer & S. Nannini, eds. *Intentionality, deliberation, and autonomy: the action-theoretic basis of practical philosophy*. Aldershot: Ashgate, pp. 17-33.
- Alfano, M., Beebe, J. & Robinson, B., (in press) The centrality of belief and reflection in Knobe effect cases: a unified account of the data.
- Alicke, M. D., 2008. Blaming badly, *Journal of Cognition and Culture* 8, pp. 179-186.
- Axelrod, R., & Hamilton, W.D., 1981. The evolution of cooperation, *Science* 221, pp. 1390-96.
- Axelrod, R., 1984. *The evolution of cooperation*. New York: Basic Books.
- Axelrod, R., 1986. An evolutionary approach to norms, *American Political Science Review* 80, pp. 1095-111.

- Barrett, H.C., 1999. Guilty minds: how perceived intent, incentive and ability to cheat influence social contract reasoning. *11th Annual Meeting of the Human Behavior and Evolution Society*. Salt Lake City, Utah.
- Beebe, J.R., (in press). Attributions of knowledge and blame in Knobe effect cases.
- Beebe, J.R., & Buckwalter, W., 2010. The epistemic side-effect effect, *Mind & Language* 25, pp. 474-98.
- Beebe, J.R., & Jensen, M., (in press). Surprising connections between knowledge and action: the robustness of the epistemic side-effect effect, *Philosophical Psychology*.
- Beebe, J.R., & Shea, J., (in press). Gettierized Knobe effects.
- Boyd, R., & Richerson, P.J., 1992. Punishment allows the evolution of cooperation (or anything else) in sizable groups, *Ethology and Sociobiology* 13, pp. 171-95.
- Buckwalter, W., & Stich, S., (in press). Gender and philosophical intuition.
- Clutton-Brock, T.H., & Parker, G.A., 1995. Punishment in animal societies, *Nature* 373, pp. 209-16.
- Cosmides, L., 1989. The logic of social exchange: has natural selection shaped how humans reason? Studies with the Wason selection task, *Cognition* 31, pp. 187-276.
- Cosmides, L., Barrett, H.C., & Tooby, J., 2010. Adaptive specializations, social exchange, and the evolution of human intelligence, *Proceedings of the National Academy of Sciences* 107, pp. 9007-14.
- Cosmides, L., & Tooby, J., 1989. Evolutionary psychology and the generation of culture, part II. Case study: a computational theory of social exchange, *Ethology & Sociobiology* 10, pp. 51-97.

- Cosmides, L., & Tooby, J., 1997. Dissecting the computational architecture of social inference mechanisms. In: G. Bock & G. Cardew, eds. *Characterizing human psychological adaptations*. Chichester: John Wiley & Sons, pp. 132-161.
- Cosmides, L., & Tooby, J., 2005. Social exchange: the evolutionary design of a neurocognitive system. In: M.S. Gazzaniga, ed. *The new cognitive neurosciences, III*. Cambridge, MA: MIT Press, pp. 1295-308.
- Craig, E., 1986. The practical explication of knowledge, *Proceedings of the Aristotelian Society* 87, pp. 211-26.
- Craig, E., 1990. *Knowledge and the state of nature: an essay in conceptual synthesis*. Oxford: Clarendon Press.
- Craig, E., 2007. Genealogies and the state of nature. In: A. Thomas, ed. *Bernard Williams*. Cambridge: Cambridge University Press, pp. 181-200.
- Evans, J., & Frankish, K., 2009. *In two minds: dual processes and beyond*. Oxford: Oxford University Press.
- Falk, A., Fehr, E., & Fischbacher, U., 2002. Testing theories of fairness and reciprocity—intentions matter. Zürich: University of Zürich.
- Fehr, E., & Gächter, S., 2002. Altruistic punishment in humans, *Nature* 415, pp.137-40.
- Fiddick, L., 1998. The deal and the danger and the evolutionary analysis of deontic reasoning. Ph.D. University of California at Santa Barbara.
- Gelfert, A., 2011. Steps to an ecology of knowledge: continuity and change in the genealogy of knowledge, *Episteme* 8, pp. 67-82.
- Gigerenzer, G., & Hug, K., 1992. Domain specific reasoning: social contracts, cheating, and perspective change, *Cognition* 43, pp. 127-71.

- Hawthorne, J., 2004. *Knowledge and lotteries*. New York: Oxford University Press.
- Hawthorne, J., & Stanley, J., 2008: Knowledge and action, *Journal of Philosophy* 105, pp. 571-90.
- Holton, R., 2010. Norms and the Knobe effect, *Analysis* 70, pp. 417-24.
- Kappel, K., 2010. On saying that someone knows: themes from Craig. In A. Haddock, A. Millar, & D. Pritchard, eds. *Social epistemology*. Oxford: Oxford University Press, pp. 69-88.
- Kelp, C., 2011. What's the point of 'knowledge' anyway?, *Episteme* 8, pp. 53-66.
- Knobe, J., 2003a. Intentional action and side effects in ordinary language, *Analysis* 63, pp. 190-3.
- Knobe, J., 2003b. Intentional action in folk psychology: an experimental investigation, *Philosophical Psychology* 16, pp. 309-24.
- Knobe, J., 2004a. Folk psychology and folk morality: response to critics, *Journal of Theoretical and Philosophical Psychology* 24, pp. 270-9.
- Knobe, J., 2004b. Intention, intentional action and moral considerations, *Analysis* 64, pp. 181-7.
- Knobe, J., 2006. The concept of intentional action: a case study in the uses of folk psychology, *Philosophical Studies* 130, pp. 203-31.
- Knobe, J., 2007. Reason explanation in folk psychology, *Midwest Studies in Philosophy* 31, pp. 90-106.
- Knobe, J., 2010. Person as scientist, person as moralist, *Behavioral and Brain Sciences* 33, pp. 315-29.
- Knobe, J., & Mendlow, G., 2004. The good, the bad, and the blameworthy: understanding the role of evaluative considerations in folk psychology, *Journal of Theoretical and Philosophical Psychology* 24, pp. 252-58.

- Kusch, M., 2009. Testimony and the value of knowledge. In: A. Haddock, A. Millar, & D. Pritchard, eds. *Epistemic value*. Oxford: Oxford University Press, pp. 60-94.
- Kusch, M., 2011. Knowledge and certainties in the epistemic state of nature., *Episteme* 8, pp. 6-23.
- Leslie, A.M., Knobe, J., & Cohen, A., 2006. Acting intentionally and the side-effect effect: theory of mind and moral judgment, *Psychological Science* 17, pp. 421-27.
- Malle, B.F., & Knobe, J., 1997. The folk concept of intentionality, *Journal of Experimental Social Psychology* 33, pp. 101-21.
- Maynard Smith, J., 1982. *Evolution and the theory of games*. Cambridge: Cambridge University Press.
- Maynard Smith, J., & Price, G.R., 1973. The logic of animal conflict, *Nature* 246, pp. 15-8.
- Morton, A., 2001. Psychology for cooperators. In: C. Morris, ed. *Practical rationality and preferences: essays for David Gauthier*. Cambridge: Cambridge University Press, pp. 153-72.
- Nadelhoffer, T., 2004. The Butler problem revisited, *Analysis* 64(3), pp. 277-84.
- Nadelhoffer, T., 2006. Bad acts, blameworthy agents, and intentional actions: some problems for jury impartiality, *Philosophical Explorations* 9, pp. 203-20.
- Nichols, S., & Ulatowski, J., 2007. Intuitions and individual differences: the Knobe effect revisited, *Mind & Language* 22, pp. 346-65.
- Pinillos, N.Á., (in press). Knowledge, experiments and practical interests. this volume.
- Richerson, P.J., Boyd, R.T., & Henrich, J., 2003. Cultural evolution of human cooperation. In: P. Hammerstein, ed. *Genetic and cultural evolution of cooperation*. Cambridge, MA: MIT Press, pp. 357-88.

- Schaffer, J., & Knobe, J., (in press). Contrastive knowledge surveyed, *Noûs*.
- Shatz, M., Wellman, H.M., & Silber, S., 1983. The acquisition of mental verbs: a systematic investigation of the child's first reference to mental state, *Cognition* 14, pp. 301-21.
- Sober, E., & Wilson, D.S., 1998. *Unto others: the evolution and psychology of unselfish behavior*. Cambridge, MA: Harvard University Press.
- Stanley, J., 2005. *Knowledge and practical interests*. New York: Oxford University Press.
- Trivers, R.L., 1971. The evolution of reciprocal altruism, *Quarterly Review of Biology* 46, pp. 35-57.
- Uttich, K., & Lombrozo, T., 2010. Norms inform mental state ascriptions: a rational explanation for the side-effect effect, *Cognition* 116, pp. 87-100.
- Williams, B., 2002. *Truth and truthfulness: an essay in genealogy*. Princeton: Princeton University Press.
- Williamson, T., 1996. Knowing and asserting, *Philosophical Review* 105, pp. 489-523.