The Proper Province of Philosophy: Conceptual Analysis and Empirical Investigation[1]

Justin Sytsma

The practice of conceptual analysis has undergone a revival in recent years. Although the extent of its role in philosophy is controversial, many now accept that conceptual analysis has at least some role to play. Granting this, I consider the relevance of empirical investigation to conceptual analysis. I do so by contrasting an extreme position (anti-empirical conceptual analysis) with a more moderate position (non-empirical conceptual analysis). I argue that anti-empirical conceptual analysis is not a viable position because it has no means for resolving conceptual disputes that arise between seemingly competent speakers of the language. This is illustrated by considering one such dispute that has been pressed by a prominent advocate of anti-empirical conceptual analysis: Bennett and Hacker (2003) assert that psychological predicates only logically apply to whole living animals, but many scientists and philosophers use the terms more broadly. I argue that to resolve such disputes we need to empirically investigate the common understanding of the terms at issue. I then show how this can be done by presenting the results of three studies concerning the application of "calculates" to computers.

Conceptual analysis as a topic of philosophical discussion has seen something of a resurgence over the past decade. Much of this resurgence owes to Frank Jackson's (1998) defense of the practice. He argues that conceptual analysis is both more important and more common than has recently been thought (see also Jackson (1994), Bealer (1987, 1998), Lewis (1994); see Laurence and Margolis (2003) for a critical discussion). Defenses like Jackson's raise a host of questions about just how important conceptual analysis is to philosophy and how frequently it should be employed. Accepting that it has at least some role to play, however, I want to focus on a subsequent question about its application: What role, if any, should empirical investigation play in conceptual analysis? Specifically, I will consider the role of empirical

[1] To appear in *Review of Philosophy and Psychology*. This research was assisted by a Dissertation Completion Fellowship, which is part of the Andrew W. Mellon / American Council of Learned Societies Early Career Fellowship Program. The author would like to thank Peter Machamer (many of the ideas in this article emerged in discussions with him over the course of writing our (2005)), Edouard Machery, Jonathan Livengood, Peter Gildenhuys, the audience at the 1st annual Interdisciplinary Approach to Philosophical Issues Conference, an anonymous referee for the *Review of Philosophy and Psychology*, and the editors of this special issue for their insightful comments and suggestions; he would also like to thank Jonathan Livengood for his assistance with the logistic regression in Section 5.2 and Mark Phelan for suggesting that experiment.

investigation in laying out the primary meaning[2] of a term for a concept that occurs in a conceptual analysis.

Focusing on the analysis of ordinary concepts (non-technical concepts), there are any number of positions that one could defend with regard to this question. At one extreme, you could argue that conceptual analysis should *always* be supported by empirical research, potentially requiring philosophers to carry out the relevant experiments themselves if the existing literature is silent on the question (think of this as experimental philosophy run wild); more moderately, you could hold that it is *good practice* to offer empirical evidence in support of relatively important or contentious bits of conceptual analysis, preferably at the time you employ them. At the other extreme, you could assert that empirical research has no role to play in conceptual analysis, arguing that it is essentially *irrelevant* to the practice. Call this *anti-empirical conceptual analysis* (AECA). More moderately, you could allow that empirical research has some role to play in conceptual analysis, but hold that it is *not generally needed* and deny that it is incumbent on the practitioner to conduct an objective investigation. Call this *non-empirical conceptual analysis* (NECA).[3]

In this paper I investigate the divide between NECA and AECA by considering a proponent of each position. I will treat Frank Jackson (1998) as an advocate of NECA. While Jackson is open to the use of empirical work in conceptual analysis, he does not think that it is typically needed. In contrast, Max Bennett and Peter Hacker (2003) seem to support AECA,

[2] As distinguished from secondary or non-literal meanings, and bearing no relation to what David Chalmers has called "primary intentions." This is discussed further in Section 1. I take the advocates of conceptual analysis that we will consider to hold that the terms they are interested in (philosophically interesting common terms) each have a primary meaning and that these are reflected in the standard employment of those terms in the general linguistic community (for our purposes, English speakers). For the sake of argument I will assume that this is correct and that the relevant terms do have such primary meanings.

[3] This list is, of course, not exhaustive of the positions one might hold on the issue; rather, it marks a few important points on what might be thought of as a spectrum—from empirical investigation being essential to conceptual analysis having no role to play.

denying that empirical work has any role to play in conceptual analysis. My primary goal is to show that in denying the relevance of empirical work to conceptual analysis, AECA has no compelling way to adjudicate between diverging assertions, made by seemingly competent speakers of the language, about the meaning of a term occurring in a conceptual analysis. I hold that this is a rather serious shortcoming and illustrate the point by considering an actual example of such a dispute, namely, the disagreement over the use of psychological predicates that is at the center of Bennett and Hacker's (2003, 2007) critique of language use in modern neuroscience. This dispute concerns two plausible hypotheses about the primary meanings of psychological predicates: Bennett and Hacker assert that these terms are restricted to whole living animals, but the scientists and philosophers that Bennett and Hacker critique do not abide by this restriction. It seems that one of these parties must be mistaken in their assessment of the primary meanings of these terms (assuming with Bennett and Hacker that the second party intends to be using these terms in accord with their primary meanings).

The question that will concern us is how we can resolve such disputes. As proponents of both NECA and AECA hold that the primary meanings of common terms can be elicited from how they are understood in the general linguistic community, it seems that the obvious way to resolve such disputes is to go out into the community and empirically investigate how people understand the terms. In fact, I hold that this is the only convincing way to adjudicate between opposed statements of the primary meaning of a common term given by two speakers of the language that otherwise seem to be competent. To illustrate, I present the results of three studies assessing one of the specific claims made by Bennett and Hacker (namely, that computers cannot literally be said to calculate).

Here is how I will proceed. In Section 1, I articulate the conception of conceptual analysis that I will be working with and detail the position of NECA as found in Jackson (1998). In Section 2, I articulate the contrasting position of AECA, focusing on the reasons that Bennett and Hacker (2003) give for denying that empirical inquiry is relevant to the practice of conceptual analysis. In Section 3, I raise the problem of conceptual disputes for AECA. The problem is illustrated via Bennett and Hacker's claim that psychological predicates only logically apply to whole living animals. In Section 4, I briefly discuss a range of empirical studies concerning the folk psychology of consciousness that speak against Bennett and Hacker's assessment of the primary meanings of some psychological predicates. In Section 5, I then present the results of the three studies that I conducted to test the application of "calculate" to computers. I argue that these studies are not only compelling evidence against Bennett and Hacker, but show how careful experimental work can be used to resolve disputes that arise in the practice of conceptual analysis. I conclude that AECA is not a viable position and draw a lesson from this for NECA.

1. Conceptual Analysis

There is not perfect agreement about what the practice amounts to between the advocates of conceptual analysis that I will consider. Nonetheless, I find that a few central themes are largely shared and I will focus on these in the discussion that follows. In particular, conceptual analysis generally concerns articulations of the primary meanings of common terms that are of philosophical interest. These meanings are thought to reflect how the terms are typically understood in the general linguistic community, leading to a concern with how they are ordinarily used. Thus, Max Bennett and Peter Hacker (2003, 6), for example, are "concerned

with the misuse of old, non-technical concepts"; they are concerned with a term for a common or garden concept being used "contrary to its ordinary use." They hold that nonsense often results from such misuse, writing:

> How can one investigate the bounds of sense? Only by examining the use of words. Nonsense is often generated when an expression is used contrary to the rules for its use. The expression in question may be an ordinary, non-technical expression, in which case the rules for its use can be elicited from its standard employment and received explanations of its meaning. (6)[4]

Following the Wittgensteinian slogan that meaning is use, Bennett and Hacker hold that the primary meaning of a common term is found in how that term is used in the general linguistic community.

Frank Jackson is not so concerned with identifying the misuse of terms and instead motivates conceptual analysis via "the importance of defining one's subject" (1998, 30). Nonetheless, this also leads to a concern with what he calls the "ordinary conception" of a given common term (this is essentially its primary meaning). Jackson writes:

> What then are the interesting philosophical questions that we are seeking to address when we debate the existence of free action and its compatibility with determinism, or about eliminativism concerning intentional psychology? What we are seeking to address is whether free action *according to our ordinary conception*, or something suitably close to our ordinary conception, exists and is compatible with determinism, and whether intentional states *according to our ordinary conception*, or something suitably close to it, will survive what cognitive science reveals about the operations of our brains. (31)

The scope of the "our" in this passage is not just philosophers, but the linguistic community more generally. Thus, Jackson equates our ordinary conception of free will, for example, with the folk theory of it (32). His goal is to articulate the folk theory and he holds that this can be done by considering people's intuitions about the applicability of the term for the folk concept in

[4] My primary concern will be with the claim, detailed in the following section, that we can examine the use of common terms without empirically investigating how those terms are employed in the linguistic community. While Bennett and Hacker might seem to suggest such an investigation in this passage, they actually hold that there is no need for competent speakers of the language to examine how other people in the linguistic community use those terms.

various possible cases. In this way, like Bennett and Hacker, Jackson is concerned with how people use the relevant terms, drawing this out by considering different scenarios.

Jackson's focus on intuitions about the applicability of a term in various possible cases is typical in the practice of conceptual analysis. Antti Kauppinen terms these intuitions, *conceptual intuitions*; these are defined (roughly speaking) as our "pre-theoretical dispositions to apply concepts to some particular cases or scenarios and to refuse to apply them to others" (2007, 96). Accepting that the primary meanings of philosophically interesting common terms are reflected in the shared conceptual intuitions of members of the linguistic community, this gives us a succinct way to distinguish between the positions of AECA and NECA. We can distinguish between them in terms of the extent of the role that their advocates assign to empirical investigation for purposes of showing that one's own conceptual intuitions are widely shared.[5] While the advocate of AECA holds that there is no need to empirically check whether their conceptual intuitions are widely shared, the advocate of NECA merely holds that empirical investigation is not typically needed.

Thus, our primary example of NECA, Frank Jackson, holds that it is fairly safe for a philosopher to generalize from her own conceptual intuitions. He holds that his intuitions, for example, generally coincide with the folk's and thereby reveal the folk theory (1998, 32). Although Jackson does not feel that it is generally necessary to conduct empirical work to confirm that his conceptual intuitions are typical, he accepts that this is an empirical question:

> I am sometimes asked—in a tone that suggests that the question is a major objection—why, if conceptual analysis is concerned to elucidate what governs our classificatory practice, don't I advocate doing serious opinion polls on people's responses to various cases? My answer is that I do—when it is necessary. Everyone who presents the Gettier cases to a class of students is doing their own bit of fieldwork, and we all know the

[5] Just how widely shared a conceptual intuition must be is not clear in the literature; and, while this is an interesting and important question, it is not one that I can do justice to here.

> answer they get in the vast majority of cases. But it is also true that often we know that our own case is typical and so can generalize from it to others. (36-37)

Jackson suggests not only that empirical work on conceptual intuitions is seldom needed, but also implies that when it is needed, rather informal methods will suffice. Nonetheless, Jackson (and the advocates of NECA more generally) does not deny that empirical investigation has a role to play in conceptual analysis. In particular, we might reasonably infer that Jackson holds that one of the times when it will be necessary to employ serious opinion polls (and presumably actual experiments if called for), is when philosophical disputes arise concerning the primary meaning of a term occurring in a given employment of conceptual analysis.

2. Anti-Empirical Conceptual Analysis

Other advocates of conceptual analysis are even less open to the use of empirical methods to check whether one's own conceptual intuitions are representative of those of the rest of the linguistic community. For example, Antti Kauppinen (2007, 95) both denies that the conceptual intuitions of the community can be "tested with methods of positivist social science" and asserts that philosophers are "entitled to appeal to intuitions about folk concepts in virtue of possessing implicit normative knowledge acquired through reflective participation in everyday linguistic practices."

Our primary example for AECA, Max Bennett and Peter Hacker, suggest an even sharper divide between the empirical and the conceptual. They begin their critique of language use in modern neuroscience by separating the empirical questions that occupy neuroscientists from the conceptual questions that philosophers are concerned with:

> Empirical questions about the nervous system are the province of neuroscience….
> By contrast, conceptual questions (concerning, for example, the concepts of mind or memory, thought or imagination), the description of the logical relations between

> concepts (such as between the concepts of perception and sensation, or the concepts of consciousness and self-consciousness), and the examination of the structural relationships between distinct conceptual fields (such as between the psychological and the neural, or the mental and the behavioral) are the proper province of philosophy. (2003, 1)

Bennett and Hacker then argue that the conceptual questions come prior to the empirical questions, writing:

> Conceptual questions antecede matters of truth and falsehood…. They determine not what is empirically true or false, but rather what does or does not make sense. Hence conceptual questions are not amenable to scientific investigation and experimentation or to scientific theorizing. For the concepts and conceptual relationships in question are *presupposed* by any such investigations and theorizing. (2003, 2)

The claim that conceptual questions are not amenable to scientific investigation is a rather radical claim. Further, it is not clear why we should accept that a given concept cannot be the target of a scientific investigation. Specifically, it is unclear why the concept *in question* must be *presupposed* by such an investigation. Certainly the investigators will have to call on a host of other concepts in carrying out any scientific investigation; but, it does not follow that they must call on their understanding of the term in question to investigate the use of that term in the linguistic community.

Elsewhere, Bennett and Hacker seem to recognize this and instead endorse a more moderate claim. They suggest that it is not so much that concepts *cannot* be investigated empirically, but that such investigations are *not needed*. Consider their (2007) response to a suggestion by Daniel Dennett (2007):

> Professor Dennett suggests that to examine the use of words involves either a form of anthropology or a form of "autoanthropology." For one has to discover the uses of words by doing appropriate social surveys, asking people to consult their intuitions on correct word usage. Alternatively, one has to consult one's own intuitions; but then it might turn out that one's intuitions diverge from those of others….
>
> This is a misconception. A competent speaker of the language no more has to consult his intuitions (hunches, guesses) than a competent mathematician has to consult his intuitions

> concerning the multiplication tables or a competent chess player has to consult his intuitions about the movements of chess pieces. (146–147)

It is worth noting that Bennett and Hacker appear to be using the term "intuition" in a different sense than Dennett (and the sense employed in this article). In equating intuitions with hunches or guesses, they suggest a clear lack of certainty about an intuition on the part of the person who has it; but, Dennett is not suggesting that people feel uncertain about the meaning of common terms. Rather, he uses "intuition" in the sense of an unreflective intellectual seeming and, in this sense the mathematician is reasonably taken to be calling on an intuition when giving the answer to a simple multiplication problem. Further, in this case the intuition is readily justified and the correct answer will be obvious to anyone with basic competence in mathematics. The issue is whether the primary meanings of common terms can be similarly justified and are equally obvious. Bennett and Hacker seem to be claiming that this is the case for "competent speakers." They suggest that while the meaning of a given term is an empirical question, it is a trivial one and one that a competent speaker has no need to engage with.

This claim critically rests on there being such "competent speakers" of the language to call on and our ability to recognize them as such (if we wish to avoid disputes). Unfortunately, Bennett and Hacker do not tell us much about how to do this. What they say is that "a competent speaker is one who has mastered the usage of the common expressions of the language" (2007, 147). But, this suggests that we would need to know the correct use of these expressions in the first place if we wanted to establish that someone was actually a competent speaker of the language. In the absence of some non-question-begging way to establish who has truly mastered the usage of the common expressions of the language, however, it seems that we would need to first conduct some empirical work to determine the primary meanings of these expressions—perhaps conducting the appropriate social surveys as Dennett suggests. This is most evident

when conceptual disputes arise between people that we would otherwise assume were competent speakers of the language.

3. Conceptual Disputes between Seemingly Competent Speakers

I hold that AECA is not a viable position, not because it cannot handle the rudder when the philosophical waters are smooth, but because it offers no compelling way of overcoming the waves of disagreement that occur. The problem is that AECA does not tell us what to do when two seemingly competent speakers have divergent conceptual intuitions about a philosophically interesting common term. Of course, one might hold that such disagreements could not occur (as Bennett and Hacker seem to suggest in responding to Dennett) or that if they did that the disagreements would be superficial could be resolved through further dialogue and reflection. But, what if we were to find ongoing disagreement between seemingly competent speakers about the primary meaning of a range of terms?

I hold that AECA has no way to convincingly resolve such disputes between seemingly competent speakers of the language exactly because it denies that empirical investigation into how those terms are used has a role to play in conceptual analysis. The reason is that the available evidence—the disputants' own conceptual intuitions—is insufficient to decide between them. Thus, it seems that for such a dispute to arise there must either be different primary meanings of a term (perhaps reflecting that we have lumped linguistic communities together that are best treated separately) or that one party in the dispute must be mistaken (and hence not actually a competent speaker of the language on Bennett and Hacker's understanding of the phrase). Deciding which of these is the case will be critical for the conceptual analysis; but, the disputants' own conceptual intuitions are clearly insufficient evidence for making a compelling

decision between these possibilities, as each intuition is countered by the other. Since it is the representativeness of the conceptual intuitions that is at issue, however, it seems that a compelling decision here will require moving beyond the disputants own intuitions and finding a standard by which to adjudicate between them. As such, I find that if such disputes occur, the advocate of conceptual analysis has little choice but to turn to an investigation of how members of the linguistic community actually use the common terms at issue.

Disputes over the primary meaning of ordinary terms do occur. In fact, at the heart of Bennett and Hacker's critique of neuroscience is just such a dispute. They argue that psychological predicates only logically apply to (some) whole animals. But, of course, other philosophers disagree (I will take myself as an example of this); further, Bennett and Hacker's critique itself indicates that many brain scientists disagree. Of course, aside from this dispute we do not have any reason to question the conceptual intuitions of these scientists or to doubt that these exceedingly well-educated researchers are competent speakers of the language. As such, the conceptual intuitions on either side of the dispute would seem to have equal standing at the outset; but, as they generate radically divergent conceptual analyses, we need some (non-question-begging) way to decide between them.

3.1 The Animal Restriction for Psychological Predicates

Although Bennett and Hacker express a general unease with the use of psychological terms in contemporary neuroscience, their critique focuses on one supposed condition of application, the animal restriction noted above. They hold that psychological predicates—such as thinking, knowing, believing, seeing (2003, 71)—only logically apply to some whole living animals.

When this restriction is violated by applying a psychological predicate to a *part* of an animal, the user commits what Bennett and Hacker term "the mereological fallacy."

As the animal restriction is the basis for this so-called fallacy, it is important to be very clear about the nature of this restriction. Two points are especially relevant. First, Bennett and Hacker are not simply saying that psychological predicates are *primarily* applied to living persons and some non-human animals, although we could perhaps discover other entities that they also apply to. Rather, they assert that it is part of the meaning of psychological predicates that they *only* apply to whole living animals. In other words, they hold that we could no more discover that a non-animal possessed one of these abilities, than we could discover a married bachelor. Second, the animal restriction is violated by the application of psychological predicates to parts of animals (resulting in the so-called mereological fallacy), but it is also violated by the application of psychological predicates to other non-animals. This is important to note because the empirical studies that I review in Section 4 and present in Section 5 concern entities that are neither animals nor parts of animals. Thus, while Bennett and Hacker primarily focus on violations of the animal restriction that are instances of the so-called mereological fallacy, the restriction is wider than this—as they realize (2007, 139).

What reasons do Bennett and Hacker offer for thinking that the primary meaning of psychological predicates includes the animal restriction? As should be expected from their commitment to AECA, they do not offer much. This restriction is simply supposed to follow from our conceptual intuitions about various psychological predicates as competent speakers of the language. Nonetheless, Bennett and Hacker do draw the restriction out by considering §281 of Wittgenstein's *Philosophical Investigations* (1958). Bennett and Hacker write:

> Wittgenstein made a profound remark that bears directly on our concerns. "Only of a human being and what resembles (behaves like) a living human being can one say: It has

> sensations; it sees, is blind; hears, is deaf; is conscious or unconscious." This epitomizes the conclusions we shall reach in our investigation. (2003, 71)

As it stands, Wittgenstein's dictum (2007, 133) diverges from Bennett and Hacker's animal restriction in a number of ways. In particular, Wittgenstein does not specify *non-human animals* as the only entities that might *behave like* human beings and he does not claim that his insight extends to all psychological predicates. As such, it is possible to both affirm Wittgenstein's dictum and deny Bennett and Hacker's animal restriction.[6]

Again, it seems that the case for the animal restriction rests on Bennett and Hacker's conceptual intuitions about psychological predicates; but, their own critique illustrates that the conceptual intuitions of seemingly competent speakers of the language are not uniform here. We therefore have a clear conceptual dispute and have been given no way to resolve it. Given that the dispute concerns how widely held the different conceptual intuitions are in the linguistic community, it seems that the way to resolve the dispute is to conduct an empirical investigation. I illustrate how this can be done in Section 5 by considering one supposed violation of the animal restriction—the application of "calculate" to computers.

3.2 The Application of "calculate" to Computers

Consider the following claim: "'The computer calculates' means no more than 'The computer goes through the electricomechanical processes necessary to produce the results of a calculation without any calculation'" (Bennett and Hacker, 2007, 152). Why should we think this? More specifically, why should we think that the literal application of "calculate" requires anything more than that a system goes through processes that reliably produce the results of calculations? Bennett and Hacker do not say. Rather, they assert that such language is metaphorical (or

[6] This is essentially what Daniel Dennett (2007, 78) does, for example; see also Machamer and Sytsma (2005), Sytsma (2007).

perhaps a dead metaphor (152), although it is not clear how this would support their basic point[7]) and insist that true calculation involves a host of *other* psychological abilities that they claim computers cannot possess. Bennett and Hacker write:

> It is true that we do, in casual parlance, say that computers remember, that they search their memory, that they calculate and sometimes, when they take a long time, we jocularly say that they are thinking things over. But this is merely a *facon de parler*. It is not a literal application of the terms…. We use computers to produce the results of a calculation—just as we used to use a slide-rule or cylindrical mechanical calculator. Those results are produced without anyone or anything literally calculating—as is evident in the case of a slide-rule or mechanical calculator. In order literally to calculate, one must have a grasp of a wide range of concepts, follow a multitude of rules that one must know, and understand a variety of operations. Computers do not and cannot. (139)

What I want to focus on in this passage is the claim that when *we* say that a computer calculates, *we* do not mean it literally.

Bennett and Hacker apparently arrive at this conclusion by consulting their conceptual intuitions about the term, although they illustrate the point by considering other artifacts besides computers. Again, I have a different conceptual intuition: I am perfectly willing to say that "computers calculate" and I mean this literally. As such, I do not find Bennett and Hacker's analogies compelling. I do not find them compelling, in part, because there seems to be an important difference between the cases: A slide-rule on its own is unable to go through the processes that produce the results of calculations, but a computer can go through the necessary processes given the appropriate initiating input. Certainly such an ability seems to be *necessary* for calculation; the question, however, was why we should think that it is not *sufficient* for calculation—and comparisons to entities that lack this ability are irrelevant to this question.

Thus, we have a conceptual dispute, with my conceptual intuitions about computers and the psychological predicate "calculates" diverging from Bennett and Hacker's. How are we to

[7] It is plausible that the term "calculate" was once restricted to human beings and only applied to machines metaphorically. Accepting this, however, it nonetheless seems that the death of the metaphor might well signal the lifting of the animal restriction for the term (see Camp, 2006, 161).

resolve this dispute? Although Bennett and Hacker have not offered any evidence for their claim that talk of "computers calculating" is meant figuratively (outside of their conceptual intuition), this does not necessarily mean that they are mistaken. Rather, I find that they make an empirical claim and the occurrence of dispute concerning it shows us that empirical evidence is needed. Furthermore, it is a claim that is readily tested. The results of three such tests are discussed in Section 5. Together, these studies indicate that, *pace* Bennett and Hacker, non-philosophers generally hold that "calculate" can be literally applied to computers. Before turning to that work, however, I want to briefly consider some recent experimental work that while not conducted to test Bennett and Hacker's animal restriction, nonetheless suggests against it.

4. Some Empirical Work relevant to the Animal Restriction

As we saw above, Bennett and Hacker hold that the conditions of application for common expressions can be elicited from their standard employment in ordinary discourse. They also claim that one such rule for psychological predicates is that they only apply to whole living animals. If Bennett and Hacker's conceptual intuitions, as competent speakers of the language, are a reliable guide to the primary meanings of common terms, then we should expect members of the linguistic community by and large to refuse to apply psychological predicates to entities that are not whole animals. In this section I review a number of studies in the growing experimental literature on the folk psychology of consciousness that show that English speakers are generally willing to apply some psychological predicates to some non-animals.

In their pioneering paper, Joshua Knobe and Jesse Prinz (2008) investigated what types of psychological predicates people were willing to apply to group agents. In their second study they presented participants with sentences applying a psychological predicate to Acme Corporation.

Knobe and Prinz found that participants judged the sentences involving the psychological predicates "believes," "intends," "wants," "knows," and "decides" to sound natural, despite the fact that each one violates Bennett and Hacker's animal restriction.[8] These results are not a fluke and have been replicated (Sytsma and Machery, 2009). In another study, Bryce Huebner and colleagues (forthcoming) gave participants a series of sentences involving a range of group agents, asking them to rate how natural they sound on the same scale used by Knobe and Prinz. On average, participants found that each of the sentences involving an intentional psychological predicate sounded natural, giving high scores to sentences such as "Sony intends to release a new product in January to increase sales" (*M*=6.5) and "Destiny's Child wants to put on a better show tomorrow night" (*M*=6.0). A similar result was found by Adam Arico (forthcoming). Further, Arico and colleagues (under review) conducted a study to test whether people found such ascriptions to be literal; they found that even after restricting the data set to participants who performed well on figurative test sentences, most judged the sentences ascribing psychological predicates to group agents to be literally true.

It appears that people are willing to ascribe psychological predicates like "believes" and "knows" to some group agents. Furthermore, as group agents are not whole living animals, these findings suggest against Bennett and Hacker's animal restriction.

Edouard Machery and I (under review) have shown a similar result for a wholly different sort of non-animal—a simple robot. In one of the cases in our first study, participants were given a vignette describing a relatively simple robot that discriminated between three colored boxes, correctly moving the red one when so instructed, and were asked whether the robot "saw red" on a scale from 1 ("clearly no") to 7 ("clearly yes"). We found that non-philosophers were generally

[8] Participants were asked to rate each sentence on a scale from 1 ("sounds weird") to 7 ("sounds natural"). Sentences included "Acme Corporation believes that its profit margin will soon increase" (*M*= 6.1) and "Acme Corporation intends to release a new product this January" (*M*= 6.3).

willing to attribute seeing to the robot (*M*=5.2). In a previously unreported follow-up, Jonathan Livengood and I ran a variation on this probe, removing anthropomorphic language and changing the target color to blue. We again found that participants were generally willing to ascribe seeing to the robot (*M*=5.1). In another study, Huebner (in press) gave participants vignettes describing one of four systems—a human, a cyborg with a human body but a robot brain, a cyborg with a robot body but a human brain, and a robot. He found that participants were willing to ascribe beliefs to the robot (such as the belief that 2 + 2 = 4 or that triangles have three sides) and there was no significant difference in this regards between the robot case and the other three systems.

As a robot is not a whole living animal, these results further indicate that people do not feel bound by Bennett and Hacker's animal restriction. Again, we have preliminary empirical evidence that it is false that people are only willing to apply psychological predicates to whole living animals.

5. Studies on the Application of "calculate" to Computers

It is unlikely that the studies described in the previous section will convince dedicated advocates of AECA, like Bennett and Hacker, that some of their conceptual intuitions are not representative. The most likely reply is to argue that the participants surveyed in those studies understood the psychological predicates figuratively. Alternatively, the advocate of AECA might argue that participants' responses do not reflect the "robust intuitions" that are at issue and that dialogue and reflection are needed to elicit these (Kauppinen, 2007). This *might* be reasonable for some philosophical thought experiments, but it is far from clear that it is reasonable for judgments about relatively straightforward sentences like the one given above.

Regardless, I hold that careful empirical investigation can overcome these worries. I illustrate this by presenting the results of three studies conducted to test one potential violation of the animal restriction that Bennett and Hacker consider. As seen in Section 3, they hold that computers cannot literally calculate. Are they representative of the linguistic community in this? As Bennett and Hacker seem to recognize that people are willing to say that computers calculate, the primary concern is whether this is meant literally or not. To test whether people find statements like "some computers can calculate" to be literal, we might simply ask them directly. One worry with this approach is that people, on average, might be rather poor at recognizing figurative language. This can be controlled for by including a range of test sentences, both literal and figurative, with the target sentence. This is what I did in my first study.

5.1 Study 1

I gave participants a total of nine sentences, counterbalanced for order. In addition to the target sentence, they were given four non-literal and four literal sentences (two true and two false).[9] For each sentence the participant was asked to indicate both whether they thought it was true or false and whether it was literal or non-literal. Each question was answered on a 7-point scale. The first was anchored at 1 with "clearly false," at 4 with "not sure," and at 7 with "clearly true"; the second was anchored at 1 with "clearly non-literal," at 4 with "not sure," and at 7 with "clearly literal." The study was administered in a classroom setting to 33 undergraduates at the University of Pittsburgh. Two participants were removed (one because she did not clearly answer the questions, one because the biographical information was incomplete). The remaining 31 participants were 61.3% female, ranging in age from 18 to 43, and with an average age of 20.2.

[9] Sentences were derived from Wolff and Gentner (2000) and Glucksberg et al. (1982).

The mean responses and standard deviation for each of the nine sentences are shown in parentheses after the sentences below (false/true; non-literal/literal) and graphically in Figure 1.

1. "Some computers can calculate." (*M*=6.55, *SD*=1.179; *M*=6.00, *SD*=1.844)
2. "Some birds are mammals." (*M*=3.65, *SD*=2.550; *M*=5.52, *SD*=2.158)
3. "Some gifts can be returned." (*M*=6.65, *SD*=0.709; *M*=5.06, *SD*=2.250)
4. "Some bachelors are married." (*M*=3.77, *SD*=2.305; *M*=3.10, *SD*=2.300)
5. "Some windows can open." (*M*=6.52, *SD*=1.262; *M*=5.45, *SD*=2.014)
6. "Some jobs are jails." (*M*=5.77, *SD*=1.521; *M*=2.03, *SD*=1.506)
7. "Some rumors can be viruses." (*M*=6.13, *SD*=1.118; *M*=2.03, *SD*=1.871)
8. "Some libraries are goldmines." (*M*=5.97, *SD*=1.354; *M*=2.19, *SD*=2.120)
9. "Some lies can boomerang." (*M*=6.00, *SD*=1.291; *M*=2.39, *SD*=2.011)

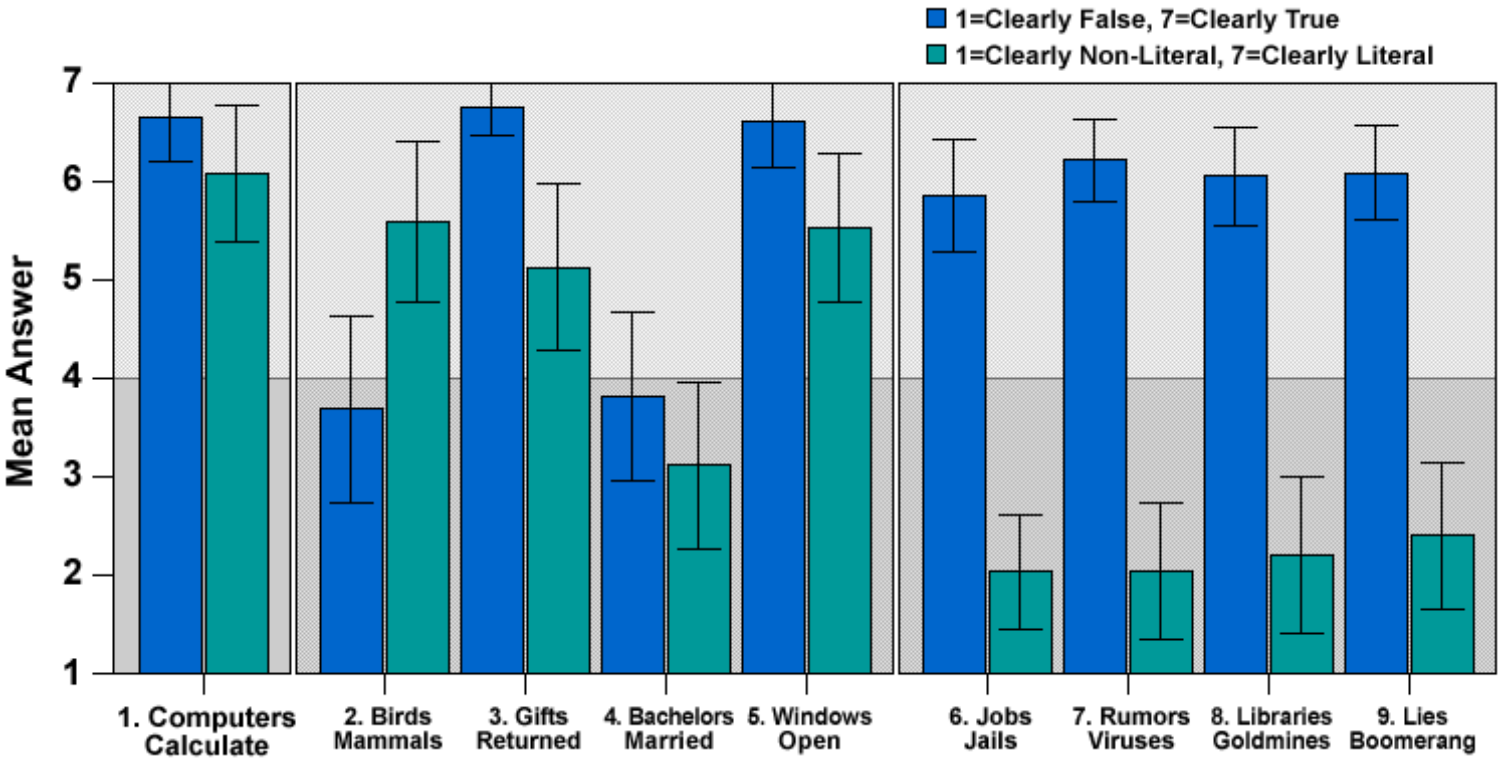

Figure 1: Study 1, results (error bars: 95% confidence intervals).

The mean responses for the eight comparison questions were largely as expected, with the exception of the false-literal sentence—"some bachelors are married"—being judged non-

literal.[10] Further, while each of the four figurative sentences were judged to be non-literal, the target sentence—"some computers can calculate"—was judged to be literal; in fact, it received the highest score for literalness out of the nine sentences. Planned analyses revealed that the mean for the literalness question for the target sentence was significantly above the neutral point of 4 ($t(30)=6.039$; $p<0.001$ (two-tailed)), significantly different from the means for each of the four figurative test questions[11], and that the means for the four figurative test questions were each significantly below the neutral point of 4[12].

Furthermore, if we restrict the data set to the 19 out of 31 participants who correctly identified the four figurative sentences as being non-literal (answering 1, 2, or 3 for each), the target sentence still has the highest score for literalness out of the nine sentences ($M=5.89$). Likewise, a planned analysis revealed that the mean for the target question remains significantly above the neutral point of 4 ($t(18)=4.256$; $p<0.001$ (two-tailed)). As seen in Figure 2, these participants clearly distinguished between the target sentence and the four figurative sentences. Further, no significant difference for the target question was seen between the means for the participants who correctly identified the figurative sentences and the participants who did not ($t(29)=0.394$; $p=0.696$ (two-tailed)).

[10] Planned analysis showed that the mean for this question was significantly below the neutral point of 4 ($t(30)=-2.186$; $p=0.037$ (two-tailed)). This might be because recognizing that the statement is literally false, participants sought a metaphorical meaning. Unlike the other literal-false sentence—"some birds are mammals"—a figurative understanding readily presents itself: As one participant indicated in the margin of the survey, some bachelors are married to their jobs.

[11] Jobs jails: $t(30)=8.897$, $p<0.001$ (two-tailed); rumors viruses: $t(30)=8.945$, $p<0.001$ (two-tailed); libraries goldmines: $t(30)=7.106$, $p<0.001$ (two-tailed); lies boomerang: $t(30)=6.950$, $p<0.001$ (two-tailed).

[12] Jobs jails: $t(30)=-7.025$, $p<0.001$ (two-tailed); rumors viruses: $t(30)=-5.587$, $p<0.001$ (two-tailed); libraries goldmines: $t(30)=-4.744$, $p<0.001$ (two-tailed); lies boomerang: $t(30)=-4.465$, $p<0.001$ (two-tailed).

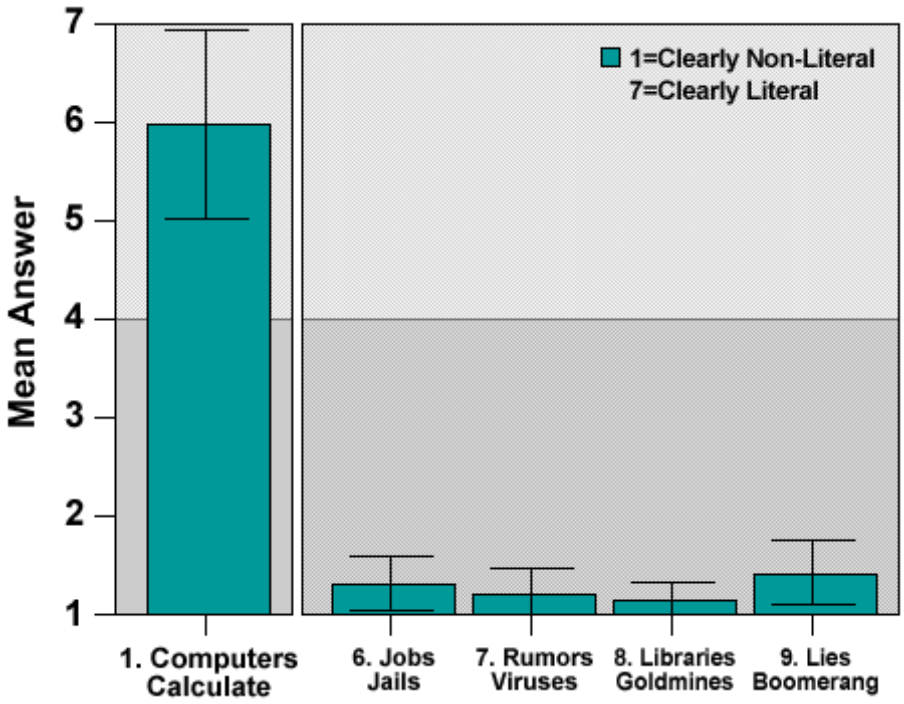

Figure 2: Literalness scores for participants correctly identifying the four figurative sentences as non-literal. (Error bars: 95% confidence intervals.)

5.2 Study 2

While the results of the first study indicate against Bennett and Hacker's contention that ascriptions of "calculate" to computers are typically meant figuratively, one might object that being able to recognize figurative language is a different skill than being able to use it. This distinction might be used to argue that the results of the first study do not reflect how participants understand the term because the study relied on participants' passive judgments about the sentence rather than actively engaging them in articulating its meaning. One way to control for this is to ask participants to paraphrase sentences that include the term "calculates." If people understand the term differently when it is applied to a computer than when it is applied to a

human (because in one case they find it to be figurative and in the other case literal), then we should expect this to be reflected in how they paraphrase the term in these divergent contexts.[13]

This is the approach taken in my second study. I gave participants one of the following two sentences and asked them to paraphrase it:

(1) "The cashier calculated John's total at the supermarket."
(2) "The cash register calculated John's total at the supermarket."

The study was administered in a classroom setting to 65 undergraduates at the University of Pittsburgh. Eight participants were removed because they did not clearly paraphrase the target term. The remaining 57 participants were 64.9% female, ranging in age from 18 to 22, and with an average age of 20.0. Of those, 27 paraphrased the first sentence, while 30 paraphrased the second. Coding the terms used for "calculated" and aggregating related variations[14], there was close agreement between the terms used for the two sentences.[15] In fact, the only term for which there was poor agreement was "rang up" and this paraphrase plausibly suggests the opposite of Bennett and Hacker's claim—it seems to suggest that it was the *cash register*, and not the

[13] This approach was suggested by Mark Phelan. See Phelan (under review) for a discussion of paraphrasing figurative language.
[14] For example, "added up" was aggregated with "added" and "added together"; while "summed" was aggregated with "summed up," "took the sum," and "produced the sum."
[15] This was tested using a logistic regression with the agent (cashier or cash register) treated as the response and the paraphrase given for "calculated" as a categorical predictor. The model was

$$\text{logit}(\pi) = \beta_0 + \beta_1 I_1 + \beta_2 I_2 + \beta_3 I_3 + \beta_4 I_4 + \varepsilon$$

where I_1 is 1 if the description was "Summed" and 0 otherwise, I_2 is 1 if the description was "Figured Out" and 0 otherwise, I_3 is 1 if the description was "Rang Up" and 0 otherwise, and I_4 is 1 if the description was "Determined" and 0 otherwise. The intercept gives the estimate for "Added Up." A likelihood ratio test accepts the hypothesis that $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$. The p-value of the test is 0.5298 (the chi-square test statistic was 3.17 with four degrees of freedom). Furthermore, the point estimate for β_0 is 0.0000 and the local p-value of the test that $\beta_0 = 0$ is 1.0000. What this means is that the odds ratio of "cashier" to "cash register" is 1:1 and none of the responses is a significant predictor. So, knowing the way a person paraphrased "calculated" does not help you to predict whether the sentence they were given was about a cashier or a cash register.

cashier, that actually calculated the total. The number of participants using each of the top five terms is shown in Figure 3.[16]

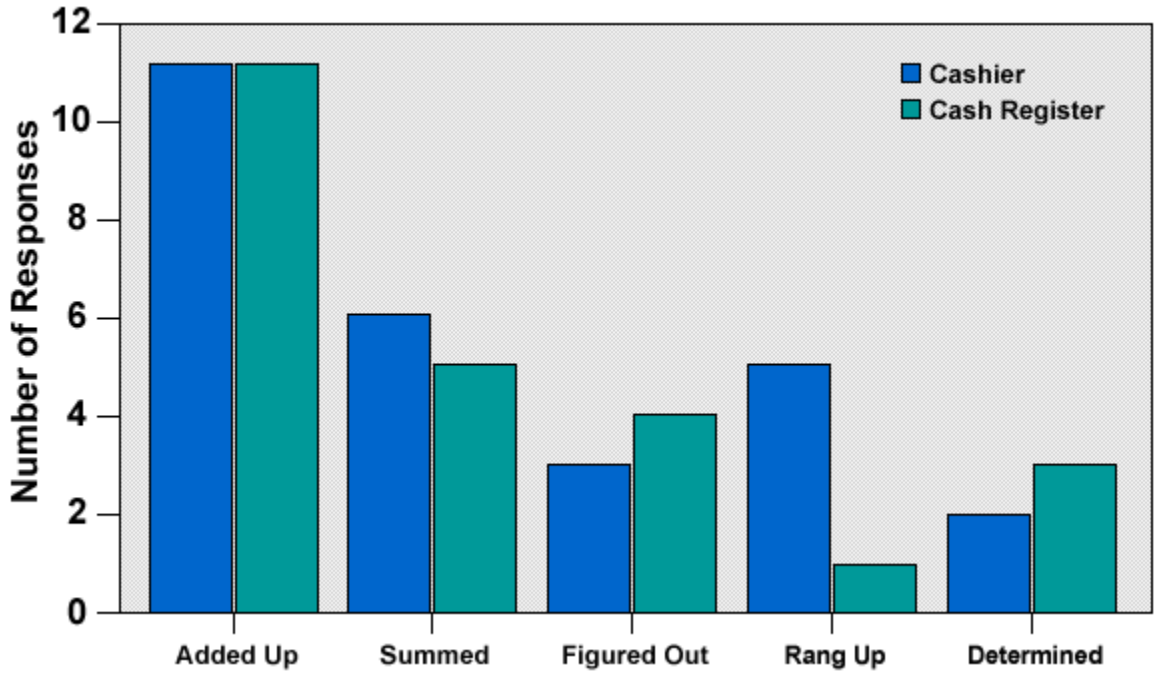

Figure 3: Study 2, results (top five answers shown).

The close agreement between the paraphrases given for "calculates" when applied to the human and the computer indicates that the participants did not understand the term differently in these two contexts. Assuming that the application of the term to the human is meant literally, this is then evidence that the application to the computer is also meant literally.

5.3 Study 3

The first two studies provide rather clear evidence that people do not generally treat the application of "calculates" to a computer as being figurative. The advocate of AECA might

[16] Additional terms not shown were "find out," "totaled," "came up with," "formulated," and "processed" with six occurrences for paraphrases of the second sentence (two for "find out" and one each for the other terms) and none for the first sentence.

nonetheless hold that these studies only elicited "surface intuitions" (as opposed to the "robust intuitions" at issue; Kauppinen, 2007) and that on further reflection (perhaps aided by philosophical dialogue) their conceptual intuitions would come to match Bennett and Hacker's. In response, one might worry that philosophical dialogue simply amounts to *training*; and, of course, people can be trained to answer such questions differently than they originally do, but this would simply show that they can be taught a *different* meaning for the term. Nonetheless, if untrained participants have robust intuitions about the application of "calculates" to computers that can be *elicited* and that diverge from their surface intuitions, then we should expect that those participants who show greater cognitive reflection should tend to give divergent answers to questions drawing on those intuitions than participants who show less cognitive reflection.

Fortunately this is something that can be tested using a standard psychological instrument—the *cognitive reflection test* (CRT) developed by Shane Frederick (2005). Frederick notes that many researchers have distinguished between those cognitive processes that are executed quickly, with little to no conscious deliberation, and those that are slower and involve deliberation and reflection. He developed the CRT as a simple measure of a person's tendency to employ the latter type of processes over the former. The test involves three problems that are relatively easy, but for which there is an intuitive but incorrect answer (each correct answer is worth 1 point, giving a scale of 0 to 3 for the test). It seems reasonable to associate the impulsive answer on these questions with what Kauppinen calls "surface intuitions"; if this is correct, then we can use CRT scores as an indicator of the likelihood that a person will reflect on a problem, potentially moving past their surface intuition if it diverges from their robust intuition.

In my third study I gave 70 participants in a classroom setting at the University of Pittsburgh the CRT along with a vignette about a computerized cash register followed by five

sentences about it. The participants were 42.9% female, ranging in age from 18 to 43, and with an average age of 21.7. For each sentence participants were asked whether it was true or false and whether it was literal or non-literal (using the same 7-point scales as in Study 1). The vignette read:

> It is an ordinary Saturday morning and Mary is at the supermarket picking up some groceries. When she went to check out there were lines in every lane that had a human cashier; several of the new "self checkout" lanes had no wait, however. In the self checkout lane, there is an item scanner attached to a cash register: You scan your items and when you are done the cash register displays how much you owe and takes your payment (having slots for cash or credit carts).
>
> Mary had never used one of these self checkout lanes before and was a bit nervous about doing so. Nonetheless, she didn't want to wait in line and decided to give it a try. Everything went fairly smoothly: Mary scanned the items in and the cash register displayed her total. It took Mary a while to find her credit card to pay, however, and after a minute the cash register started to beep. When Mary inserted her credit card, the beeping stopped. Mary finished paying with no further trouble. Overall, Mary thought the self checkout lane worked quite well and thought that she would use it again.

The mean responses and standard deviations for each of the sentences are shown in parentheses after the sentences below (false/true; non-literal/literal) and graphically in Figure 4.

1. "The cash register calculated Mary's total." (M=6.59, SD=0.940; M=6.20, SD=1.638)
2. "The cash register is a computer." (M=6.57, SD=1.149; M=5.96, SD=1.601)
3. "The cash register knew how much each item cost." (M=5.37, SD=2.310; M=3.87, SD=2.553)
4. "The cash register desires money." (M=3.47, SD=2.370; M=2.63, SD=2.107)
5. "The cash register was angry when Mary didn't pay." (M=2.51, SD=2.090; M=1.94, SD=1.895)

As in the first study, participants on average judged the sentence ascribing "calculates" to the cash register as being both true and literal. Planned analysis revealed that the mean response for the literalness of the sentence was significantly above the neutral point of 4 (t(69)=11.235, p<0.001 (two-tailed)). Further, participants judged that the application of two of the other psychological predicates to the cash register were both false and non-literal ("desires," "angry").

Interestingly, participants treated “the cash register knew how much each item cost” as true but were split on whether it was literal or not.

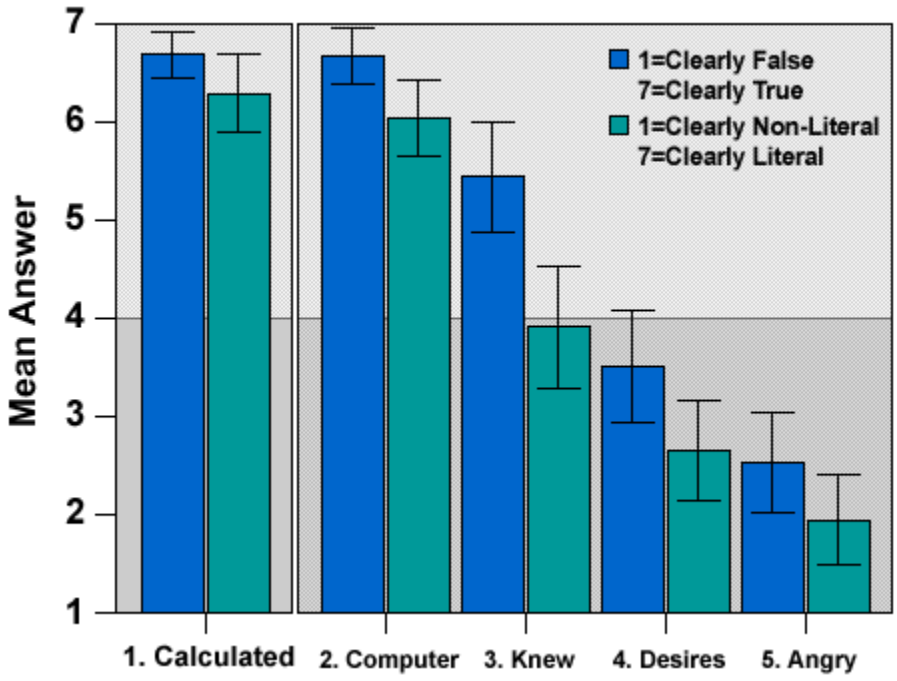

Figure 4: Study 3, results (error bars: 95% confidence intervals).

If the “robust intuitions” hypothesis were correct, then we would expect participants with higher CRT scores to be more likely to treat the ascription of “calculates” to the cash register as either false or figurative. This is not what we find, however. In fact, CRT score shows a slight positive correlation with the responses for both the truth and literalness of the target sentence (truth: $r(70)=0.056$, $p=0.648$, $r^2=0.003$; literalness: $r(70)=0.263$, $p=0.028$, $r^2=0.069$). Put another way, more reflective participants are *not* more likely to deny that a computerized cash register could literally calculate than less reflective participants.

Taken together, these three studies are compelling evidence against Bennett and Hacker’s claim that the primary meaning of “calculates” includes the animal restriction. The evidence suggests that by and large people are not only willing to apply the term to computers, but that

they mean it literally. More importantly than simply showing that Bennett and Hacker are mistaken in this case, however, these studies illustrate how empirical investigation can help resolve conceptual disputes, thereby aiding the practice of conceptual analysis.

6. Conclusion

Accepting that conceptual analysis has some role to play in modern philosophy, there are important methodological questions to be asked about its employment. I have focused here on the relevance of empirical investigation to conceptual analysis. I have argued against those thinkers who hold that empirical investigation is irrelevant to conceptual analysis, showing that it gives us a needed means for resolving the philosophical disputes that arise with regard to the primary meaning of terms for folk concepts. I illustrated this by considering one such dispute.

Having seen that empirical investigation can play a needed role in conceptual analysis, it is worth asking how widely it should be employed and how it should be carried out. These are difficult questions that I cannot answer here. Nonetheless, I find that the case against AECA suggests a moral for NECA. Specifically, it suggests that if one is going to employ conceptual analysis, then it is *good practice* to conduct formal empirical investigations of the key terms used in the analysis.

Recognizing that disputes can arise and that some seemingly competent speakers of the language have conceptual intuitions that are not representative of the intuitions of the larger linguistic community, each of us has reason to at least entertain the possibility that some of our conceptual intuitions are not representative and that reliance on them could generate non-substantive philosophical disputes. Assuming that avoiding such philosophical disputes is good practice in philosophy, then it will be good practice to conduct empirical investigations more

liberally in conceptual analysis than the advocate of NECA seems to suggest (see Section 1). Furthermore, if the goal is to avoid the unnecessary philosophical disputes that can arise in employing conceptual analysis, then I find that the rather informal methods suggested by Frank Jackson (1998, 36-37) will not suffice. The reason is that they do not control for a number of well-known biases. In particular, when Jackson suggests that simply presenting cases to a class of students is to do the necessary empirical work, he is ignoring the possibility that his philosophy students might feel pressure to express support for the response that Jackson is looking for or that his own bias might influence his perception of the assent of his students; but, the possibility of these biases render such casual fieldwork rather unconvincing. Bluntly, informal methods (such as a show of hands in an introductory philosophy class) are likely to be too informal to be compelling in cases of actual dispute. And, as the goal of the work is to resolve the dispute, this is good reason to turn to the more rigorous methods employed, for example, in experimental psychology.

References

Arico, A. Forthcoming. Folk Psychology, Consciousness, and Context Effects. *Review of Philosophy and Psychology*.

Arico, A., Fiala, B. Goldberg, R., and Nichols, S. Under review. The Folk Psychology of Consciousness.

Bealer, G. 1987. The Philosophical Limits of Scientific Essentialism. *Philosophical Perspectives*, 1 pp. 289-365.

Bealer, G. 1998. Intuition and the Autonomy of Philosophy. In Depaul, M., and Ramsey, W. (eds.), *Rethinking Intuition* (pp. 201–239). Oxford: Rowman and Littlefield.

Bennett, M. R., and Hacker, P. M. S. 2003. *Philosophical foundations of neuroscience*. Oxford: Blackwell.

Bennett, M. R., and Hacker, P. M. S. 2007. The Conceptual Presuppositions of Cognitive Neuroscience: A Reply to Critics. Bennett, M. R., Dennett, D., Hacker, P. M. S., and Searle, J., *Neuroscience and Philosophy: Brain, Mind, and Language* (pp. 127–162). New York: Columbia University Press.

Camp, E. 2006. Metaphor in the Mind: The Cognition of Metaphor. *Philosophy Compass*, 1(2) pp. 154–170.

Dennett, D. 2007. Philosophy as Naïve Anthropology: Comment on Bennett and Hacker. Bennett, M. R., Dennett, D., Hacker, P. M. S., and Searle, J., *Neuroscience and Philosophy: Brain, Mind, and Language* (pp. 73–96). New York: Columbia University Press.

Frederick, S. 2005. Cognitive Reflection and Decision Making. *Journal of Economic Perspectives*, 19(4) pp. 25–42.

Glucksberg, S., Gildea, P., and Bookin, H. 1982. On understanding nonliteral speech: Can people ignore metaphors? *Journal of Verbal Learning and Verbal Behavior*, 21 pp. 85–98.

Huebner, B. In press. Commonsense concepts of phenomenal consciousness: Does anyone care about functional zombies? *Phenomenology and Cognitive Science.*

Huebner, B., Sarkissian, H., and Bruno, M. Forthcoming. What does the Nation of China think about phenomenal states? *Review of Philosophy and Psychology*.

Jackson, F. 1994. Armchair Metaphysics. In Jackson, F. (1998), *Mind, Method, and Conditionals: Selected Essays* (pp. 154–176). London: Routledge.

Jackson, F. 1998. *From Ethics to Metaphysics*. Oxford: Oxford University Press.

Kauppinen, A. 2007. The Rise and Fall of Experimental Philosophy. *Philosophical Explorations*, 10(2) pp. 95–118.

Knobe, J., and Prinz, J. 2008. Intuitions about Consciousness: Experimental Studies. *Phenomenology and Cognitive Science,* 7(1) pp. 67–83.

Laurence, S., and Margolis, E. 2003. Concepts and Conceptual Analysis. *Philosophy and Phenomenological Research*, 67(2) pp. 253–282.

Lewis, D. 1994. Reduction in Mind. In Lewis, D. (1999), *Papers in Metaphysics and Epistemology* (pp. 291–324). Cambridge: Cambridge University Press.

Machamer, P., and Sytsma, J. 2005. Neuroscienze e natura della filosofia. *Iride*, 46 pp. 495–514.

Phelan, M. Under review. The Inadequacy of Paraphrase is the Dogma of Metaphor.

Sytsma, J. 2007. Language Police Running Amok. *Journal of Theoretical and Philosophical Psychology*, 27(1) pp. 89–203.

Sytsma, J., and Machery, E. 2009. How to study Folk Intuitions about Phenomenal Consciousness. *Philosophical Psychology*, 22(1) pp. 21–35.

Sytsma, J., and Machery, E. Under review. Two Conceptions of Subjective Experience.

Wittgenstein, L. 1958. *Philosophical Investigations, Third Edition*. Upper Saddle River, NJ: Prentice Hall.

Wolff, P., and Gentner, D. 2000. Evidence for role-neutral initial processing of metaphors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26 pp. 529–541.