The Robots of the Dawn of Experimental Philosophy of Mind

Justin Sytsma

But then, it is the obvious which is so difficult to see most of the time. People say 'It's as plain as the nose on your face.' But how much of the nose on your face can you see, unless someone holds a mirror up to you?

-- Isaac Asimov, "The Evitable Conflict," *I, Robot*

...studying minds as I do, I can tell dimly that there are laws that govern human behavior.... They may be statistical in nature, so that they might not be fruitfully expressed except when dealing with huge populations. They may be very loosely binding, so that they might not make sense unless those huge populations are unaware of the operation of those laws.

-- Isaac Asimov, *The Robots of Dawn*

0. Little Lost Robot

Philosophers of mind have often called on examples of non-humans in shaping their accounts of the mental—from Leibniz's mill and automata, to the nation of china and other group agents, to zombies and Martians, to bats and other animals, to a host of computer systems, cyborgs, androids, and robots. Experimental philosophers of mind have been no exception. For example, in their pioneering work investigating the folk theory of consciousness, Joshua Knobe and Jesse Prinz (2008) called on examples of group agents, fish, and an enchanted chair. Other contributions to the literature have called on examples ranging from God and a frog (Gray, Gray, and Wegner, 2007), to plants and insects (Arico, Fiala, Goldberg, and Nichols, 2011), to Destiny's Child and the Catholic Church (Huebner, Bruno, and Sarkissian, 2010), to monkeys and metallic slugs (Sytsma and Machery, 2012a), to ghosts and spirits (Buckwalter and Phelan, forthcoming-a), to sophisticated cyborgs and robots (Huebner, 2010), and many more examples besides.

In our response to Knobe and Prinz's article, Edouard Machery and I (Sytsma and Machery, 2009) argued for using another type of non-human example—a simple robot. This was done in a follow-up article (Sytsma and Machery, 2010) presenting data on judgments about a non-humanoid robot named Jimmy.[1] Jimmy has since appeared in experiments reported in several other articles (Buckwalter and Phelan, forthcoming-b; Sytsma and Machery, 2011; Sytsma, forthcoming), including Fiala, Arico, and Nichols's (FAN) contribution to this volume. In this chapter, I will consider what lessons we should draw from work on this little robot from

[1] I would like to say that the name "Jimmy" was selected with Asimov's "Robbie" in mind (the title character from the first story in the collection *I, Robot*). But, alas, Jimmy was based on another famous robot—Shakey, developed at the Artificial Intelligence Center of Stanford Research Institute during the second-half of the 1960s. In fact, the image used for Jimmy was selected to bear a resemblance to a slimmed-down version of Shakey.

the dawn of experimental philosophy, focusing on the experimental studies reported by FAN in the previous chapter.

I begin, in Section 1, with a discussion of the different objectives driving the work of Knobe and Prinz (2008), Sytsma and Machery (2010), and FAN, distinguishing between the positive and negative hypotheses found in these papers. In Section 2, I note that while the truth of FAN's positive hypothesis is compatible with the truth of Sytsma and Machery's negative hypothesis, our empirical findings are potentially problematic for FAN's account; in turn, the empirical findings reported by FAN in addressing this issue are potentially problematic for our negative hypothesis. I address this issue in Section 3, presenting the results of four new studies that support Sytsma and Machery's negative hypothesis against the challenge raised by FAN. Finally, in Section 4, I argue that while the results of my new studies support our negative hypothesis, they should not be seen as being otherwise problematic for FAN: FAN's positive hypothesis and Sytsma and Machery's negative hypothesis can coexist peacefully.

1. Reason

The beginnings of experimental philosophy of mind can reasonably be traced back to Knobe and Prinz's (2008) article on intuitions about consciousness. This article is focused on a *positive* objective: Knobe and Prinz open by noting that they aim to investigate people's intuitions for their own sake, asserting that they are worthy of study in their own right, beyond any implications these intuitions might have with regard to philosophical accounts of consciousness. And Knobe and Prinz profess to have no ulterior motives in conducting this inquiry.

This is not the case in our response, however—we had ulterior motives in Sytsma and Machery (2010). In fact, our first objective in that paper was *negative*: We called on the empirical data we presented to raise doubts concerning a common assumption in recent philosophical work on consciousness. Thus, we noted that the existence of phenomenally conscious mental states is often taken to be obvious from first-person experience with states like seeing red and feeling pain. We then reasoned that if this is indeed the case, then we should find that lay people tend to classify mental states in the same way that philosophers do. In particular, they should tend to treat such mental states similarly. This was tested for attributions of mental states to the simple robot Jimmy. Against the prediction derived from the philosophical tradition, we found that lay people (i.e., the "folk" or people with little to no training in philosophy or consciousness studies) do not treat seeing red and feeling pain similarly. While lay people tend to accept that Jimmy sees red, they tend to deny that Jimmy feels pain. We concluded that our findings cast doubt on the claim that the existence of phenomenally conscious mental states is obvious from first-person experience.

We did not stop there, however, but went on to pursue a positive objective as well. Having argued that lay people do not tend to classify mental states as philosophers do, we wanted to determine how they do classify mental states. We proposed that this was based not on the supposed distinction between mental states that are or are not phenomenally conscious, but on the distinction between mental states that have or lack valence (a hedonic value for the subject). I have since raised doubts about this *valence account* (Sytsma, forthcoming); and, calling on recent work by experimental philosophers on the folk theory of perception (Reuter, 2011; Sytsma, 2010; Reuter, Phillips, and Sytsma, forthcoming), I have put forward an alternative positive hypothesis—the *naive account*.[2]

[2] In brief, I call on previous findings that lay people tend to take both colors and pains to be mind-independent qualities of objects outside of the mind/brain, using this to explain the difference in responses to the two Jimmy probes in Sytsma and Machery's (2010) first study: The key difference is that while people tend to hold that red is present for Jimmy to see, they tend to deny that pain is present for Jimmy to feel.

Like Knobe and Prinz, but in contrast to Sytsma and Machery, FAN focus on a positive objective. In a fascinating series of articles, FAN have put forward what they term the *agency model* of mental state attribution (Arico, Fiala, Goldberg, and Nichols, 2011; Fiala, Arico, and Nichols, 2011; Fiala, Arico, and Nichols, present volume). According to this model, lay mental state attributions result from a dual-process cognitive system, with one of the processes taking the *low road* (operating in a fast, automatic, and domain-specific way) while the other takes the *high road* (operating in a slow, deliberate, and domain-general way). FAN then argue that low-road processing categorizes entities as AGENTs based on cues such as their having facial features, displaying interactive behavior, or moving with distinctive trajectories. Such a categorization is then thought to be sufficient to incline a person to ascribe a wide range of mental states to the AGENT, be those mental states phenomenally conscious or not, and be they valenced or not. As such, the agency model predicts that people will be inclined to ascribe mental states like feeling pain to a simple robot that displays AGENT cues, such as the robot Jimmy.[3]

Recall, however, that Sytsma and Machery (2010) found that lay people tend to deny that Jimmy feels pain. At first glance, this finding seems to pose a problem for the agency model. FAN predict that people will be disposed to attribute a wide range of mental states to an entity like Jimmy, including feelings of pain, but the responses we collected were not in line with such a disposition. FAN have a ready response, however: Dispositions can be blocked. In particular, they can argue that the judgments elicited in our studies are the result of high-road processing that has overridden the dispositions produced by the low-road processing described by the

[3] In Sytsma and Machery (2010) we describe Jimmy as both exhibiting interactive behavior and moving in a distinctive trajectory—two of FAN's AGENT cues. Further, in a follow-up study reported in Sytsma and Machery (2012a), we describe Jimmy as either having or lacking a face with changing expressions displayed on a computer monitor, adding a third AGENT cue. We found that the lay people surveyed tended to affirm that Jimmy saw blue and to deny that Jimmy felt pain, whether or not Jimmy was described as having a face.

agency model. If this is correct, then Sytsma and Machery's empirical findings do not pose a problem for FAN's agency model. And FAN present evidence that this is the case.

2. The Evitable Conflict

In responding to the potential difficulty noted in the previous section, FAN argue that the responses to Sytsma and Machery's (2010) Jimmy probes "are not wholly the product of low-road processing" (8). The reason they offer is that in studies like ours, participants "have an opportunity to spend some time engaging in conscious, high-road reflection before making their judgments about robots" (8). Of course, participants need not take advantage of the opportunities afforded them; nonetheless, it does seem reasonable to suppose that when given the chance, most will engage in at least some reflection before selecting an answer.

Further, in Sytsma and Machery (2012b) we presented empirical findings that arguably suggest that lay people do in fact employ high-road processing in responding to our Jimmy probes. These findings were given in response to a critique by Brian Talbot (2012). In that article, Talbot argues that the responses we reported in our original paper do not support our negative hypothesis because those responses were the result of low-road processing, while responses rooted in high-road processing are what is needed. We countered, in part, by presenting evidence that the same pattern of responses found in Sytsma and Machery (2010) holds when steps are taken to ensure that participants employ high-road processing. This evidence comes from three studies.

In the first study, participants were given a revised version of the Jimmy "sees" probe from our original article, correcting for a few potential issues that have been noted elsewhere, including switching the target color to blue.[4] Participants were then given a measure of how likely they are to override low-road processing and give answers that reflect high-road processing—Shane Frederick's (2005) three question Cognitive Reflection Test (CRT). As expected, we found that the majority of participants answered that Jimmy saw blue; more importantly, there was no correlation between participants, responses and their CRT score. In our second study, we gave participants the CRT before the Jimmy probe, arguing that the CRT questions would prime reflective individuals to engage high-road processing. Once again, we found that the majority of the "high CRT" participants—participants answering at least one CRT question correctly—responded that Jimmy saw blue. Finally, in our third study, we used another means of priming high-road judgments. Atler, Oppenheimer, Epley, and Eyre (2007) found that people can be induced to employ high-road processing by making the problem difficult to read. To do this we gave participants the Jimmy vignette using an extremely low-quality printout. And, yet again, we found that the majority of participants answered that Jimmy saw blue.

Setting aside worries about the adequacy of dual-process models with respect to mental state attributions, the results presented in Sytsma and Machery (2012b) still do not *necessarily* imply that participants employed high-road processing in our original study. Thus, it might be that participants in that study employed low-road processing, but that low-road processing tends to produce the same judgments about such cases as high-road processing. Nonetheless, given the reason provided by FAN for expecting that high-road judgments about robots will differ from low-road judgments, and noting that the responses of lay people remain effectively the same when we specifically attempt to elicit high-road judgments, we can tentatively conclude that the

[4] The vignette for these studies matches that given in Section 3 of this chapter. Changes from the original include changing the target color from red to blue (in previous testing we found that a few participants gave comments suggesting that they understood "sees red" metaphorically), removing anthropomorphic language, and having the test repeated three times to ease skeptical doubts.

original probes from Sytsma and Machery (2010) elicit high-road judgments. And I will assume that this is the case for the remainder of this chapter.

Taking the responses presented in Sytsma and Machery (2010) to reflect high-road processing, there is no conflict between FAN's positive hypothesis and the evidence given for our negative hypothesis, since the former makes a claim about low-road processing, while the latter reflects high-road processing. Things are not actually quite so peaceful as this might suggest, however. Thus, in their contribution to the present volume, FAN *also* argue that if high-road processing is overriding participants' low-road intuitions, then we should expect them to tend to deny that robots have mental states in general. And this prediction is at odds with our finding that lay people tend to affirm that Jimmy saw red.

In support of their prediction, FAN offer both a theoretical reason for expecting people to resist attributing mental states to robots when employing high-road processing and empirical evidence suggesting that people are less willing to attribute mental states like seeing red to Jimmy than was indicated by our previous findings. With regard to the theoretical reason for their prediction, FAN write:

> It is effectively a platitude in our culture that robots are incapable of pain or emotion. Given the cultural prevalence of that attitude, it is reasonable to hypothesize that this belief will figure in high-road reasoning about robots. If so, then subjects will show significant resistance to attributions of mental states to robots generally. (8)

This conclusion does not follow, however. Rather, accepting that it is a platitude in our culture that robots are incapable of feeling pain and emotion, it simply follows that we should expect people to resist attributing *some* mental states to robots, not mental states in general. Specifically, we should expect people to resist attributing feelings of pain and emotions to robots. But, this expectation is compatible with our original results: In Sytsma and Machery (2010) we found that people tend to deny that Jimmy felt pain and to deny that Jimmy felt anger (in our second study).

In fact, not only do our results seem to be compatible with the platitude noted by FAN, but thinking about this platitude played a role in the development of our valence account of lay mental state attributions.

While I do not find FAN's theoretical reason for expecting lay people to generally resist attributing mental states to robots to be compelling, their empirical results are another matter. And the results of their first study suggest that despite our previous findings, people are *not* generally willing to ascribe mental states of seeing to the robot Jimmy. FAN motivate their study by arguing that our participants had "no way of describing Jimmy's information-processing behavior besides adverting to mental states" (8). In other words, FAN suspect that our participants found it more informative—if ultimately inaccurate—to affirm that Jimmy saw red because this is the only way that they had to express that Jimmy did something like seeing, such as detecting the color of the box.[5]

To test their objection, FAN carried out a study in which they gave participants one or the other of the two vignettes used for "seeing" in the first study in Sytsma and Machery (2010)—either the vignette describing Jimmy or the corresponding vignette describing a normal human Timmy. The only difference between FAN's vignettes and ours is that FAN changed the target color from red to green. In addition, FAN changed the question that was asked about the vignette. They asked participants to select those statements that seemed right to them from a list of five given in the following fixed order:

> Jimmy/Timmy detected green.
> Jimmy/Timmy saw green.

[5] I do not find it to be quite so clear that participants really had no way to express that Jimmy detected red without affirming that Jimmy saw red in our original study. The first reason is that participants answered the question "Did Jimmy see red?" on a 7-point scale such that partial disagreement plausibly could be expressed by selecting a midpoint answer. Second, participants were also asked to explain their answers. And if a significant percentage of participants had affirmed that Jimmy saw red due to a desire to note Jimmy's information-processing behavior, then we would expect many of these participants to articulate this in their explanations—but that is not what we found.

Jimmy/Timmy located the green box.
Jimmy/Timmy identified the green box.
Jimmy/Timmy moved the red box.

Excluding those participants who answered that Jimmy/Timmy moved the red box, which was included in the list as a materials check[6], FAN found that only 28% (7 out of 25) selected "Jimmy saw green" compared to 57% (16 out of 28) selecting "Timmy saw green." This difference is significant. The complete results for this study are shown in Figure 1.

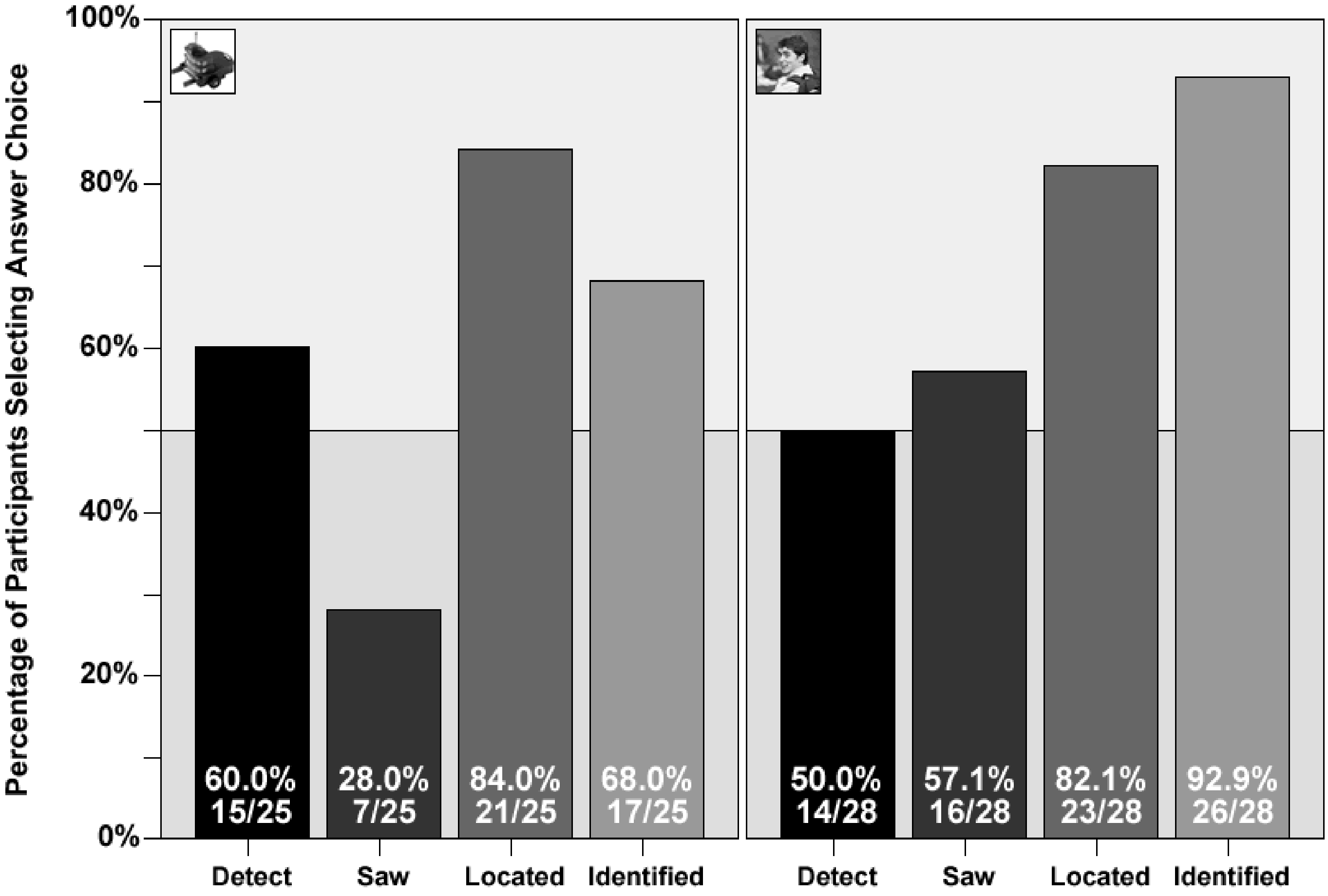

Figure 1. Results of Study 1 from Fiala, Arico, and Nichols.

[6] It might be thought that it is not so much incorrect to say that Jimmy/Timmy moved the red box, as that this is underdetermined by the vignettes: While Jimmy/Timmy is described as moving the green box, the vignettes do not preclude that the red box was moved in the process. Despite this quibble, I will follow FAN in excluding participants who answered that Jimmy/Timmy moved the red box in the new studies reported in the next section.

Based on their results, FAN conclude that by only asking participants whether Jimmy saw red in the first study in Sytsma and Machery (2010), responses were inflated. FAN then argue that their results suggest that although most people do not hold that Jimmy sees, participants want to communicate that the robot did process information concerning the color of the boxes and do this by affirming that Jimmy saw red.

While the results of FAN's new study are intriguing, I nonetheless hold that their conclusion should be resisted. The primary reason for skepticism is that other facets of FAN's results are quite puzzling, suggesting that the structure of *their* probe question is deflating participant responses rather than that the structure of our probe question inflating participant responses. Most notably, a surprisingly low percentage of participants in FAN's study answered that Timmy saw green (57%), and an even lower percentage answered that Timmy detected green (50%)! This is quite puzzling, since Timmy is described as being a *normal human* who correctly performs a visual task involving color detection. In other words, it would seem to be rather unproblematic to say both that Timmy saw and detected in this case, yet participants were not significantly more likely to select those options than chance.[7]

The low numbers for "saw" and "detect" on FAN's Timmy probe plausibly indicate that something is going awry in their study. And if something about the structure of the probe is depressing responses for those options for the Timmy probe, it is reasonable to expect that a similar effect is depressing responses for those options for the Jimmy probe. In fact, there are a number of potential issues that might be at play here. First, it is worth noting that the sample sizes for FAN's study were relatively small and that they did not restrict responses to lay people, as was done in Sytsma and Machery (2010). These factors motivate replication of their study.

[7] For detect, the proportion is exactly what one would expect by chance. For saw: One-sample test of proportion with continuity correction, χ^2=0.3214, df=1, p=0.5708.

More seriously, the sentence structure was not the same across the five answer choices in FAN's probes: While the final three answer choices specify an object (they concern "the green box"), the "saw" and "detect" answer choices do not (they simply concern "green"). But, it might be that people find it somewhat strange to say that an agent sees or detects a *property*, as opposed to an *object* with that property, especially when the distinction is made salient (as it is in FAN's probes due to the contrast between the two types of answer choices). As such, participants might shy away from answering that Jimmy/Timmy "saw green" and "detected green" in this context, even if they are willing to ascribe states of seeing and detecting to the agent. Another potential issue is that FAN presented the five answer choices in a fixed order in which the "detect" statement was given first and the "saw" statement was given second. As such, it is possible that an ordering effect is depressing answers of "saw" relative to the other answer choices. This concern would be exacerbated if participants are hesitant to select all of the first four answer choices, suspecting that at least one of them must be incorrect. For example, if participants tend to understand "detect" as being synonymous with "see" in this context—as I have argued is the case elsewhere (Sytsma, 2009; Sytsma, forthcoming)—then they might be inclined to select only one of these two answer choices, and which one they pick might be influenced by an ordering effect.

Of course, that such concerns can be raised about FAN's results does not mean that their results should be dismissed—a point I've urged repeatedly in other contexts (Sytsma and Livengood, 2011; Sytsma and Livengood, forthcoming). Instead, these concerns should simply be taken to motivate further empirical studies that attempt to control for the issues raised. I conducted four such studies, as discussed in the following section.

3. Evidence

To further investigate FAN's study, I began by attempting to replicate their results using the revised version of the Jimmy probe from Sytsma and Machery (2011). Participants were asked to carefully read one or the other of the following two vignettes:

> Jimmy (shown below) is a relatively simple robot built at a state university. Jimmy is equipped with a video camera, wheels for moving about, and two grasping arms for moving objects.
>
>
>
> As part of an experiment, Jimmy was put in a room that was empty except for one blue box, one red box, and one green box (the boxes were identical in all respects except color). Jimmy was instructed to put the blue box in front of the door. Jimmy performed the task correctly and with no noticeable difficulty. The test was then repeated on three consecutive days with the order of the boxes shuffled. Each time Jimmy correctly moved the blue box, doing so with no noticeable difficulty.
>
> Timmy (shown below) is a normal undergraduate at a state university.
>
>
>
> As part of an experiment, Timmy was put in a room that was empty except for one blue box, one red box, and one green box (the boxes were identical in all respects except color). Timmy was instructed to put the blue box in front of the door. Timmy performed the task correctly and with no noticeable difficulty. The test was then repeated on three consecutive days with the order of the boxes shuffled. Each time Timmy correctly moved the blue box, doing so with no noticeable difficulty.

Participants were then asked to select each of the statements that they agreed with from the list shown below. The answer choices used the same sentence structure and fixed ordering as in FAN's study—the only differences were that the target color was changed from green to blue and that a second materials check was added:

> Jimmy/Timmy detected blue.
> Jimmy/Timmy saw blue.
> Jimmy/Timmy located the blue box.
> Jimmy/Timmy identified the blue box.
> Jimmy/Timmy moved the blue box.
> Jimmy/Timmy moved the red box.

Responses were collected from 90 participants who passed the materials checks (answered that Jimmy/Timmy moved the blue box and did not answer that Jimmy/Timmy moved the red box).[8] The results are shown in Figure 2.

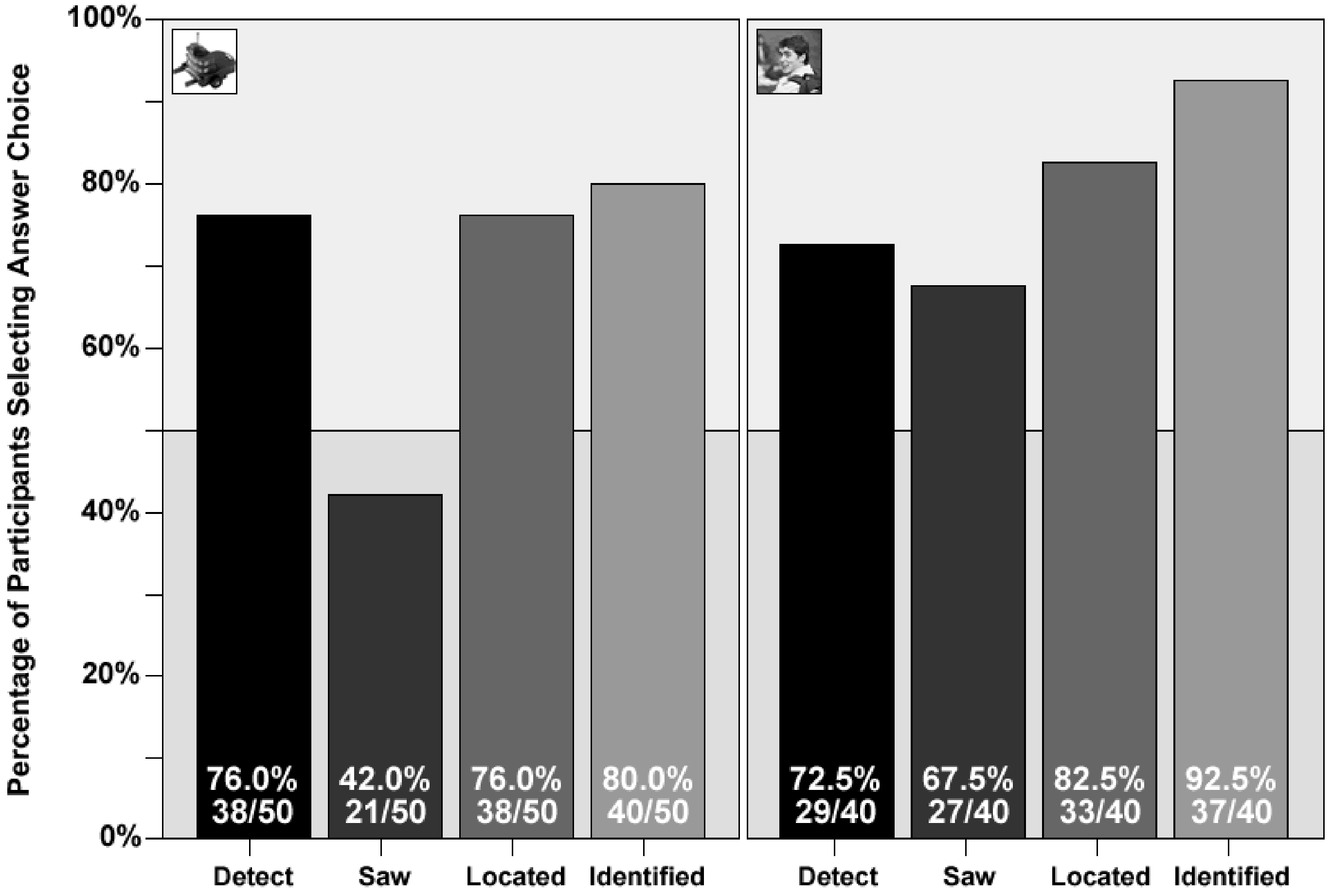

Figure 2. Results of Study 1: Replication of Study 1 from Fiala, Arico, and Nichols.

The first thing to note about the results of my first study is that, as expected, I found a rise both in the percentage of participants answering that Timmy saw (67.5% versus 57.1%) and

[8] In each of the four studies reported in this section, participants were native English-speakers, 18 years of age or older, with at most minimal training in philosophy (excluding philosophy majors and those who have taken graduate-level courses in philosophy). All participants were recruited through the Philosophical Personality website (http://philosophicalpersonality.com).

in the percentage answering that Timmy detected (72.5% versus 50.0%).[9] More importantly, I found an increase in the percentage of participants answering that Jimmy saw: While only 28.0% of Fan's participants "Jimmy saw green," 42.0% of my participants selected that "Jimmy saw blue."[10] In fact, while it remains the case that fewer than half of the participants answered that Jimmy saw, the percentage is no longer significantly below the 50% mark.[11]

The results of my first study suggest that things are perhaps not quite so dire for the conclusion that lay people tend to hold that Jimmy sees as FAN's results suggest. To further test this conclusion, in my second study I changed the first two answer choices from my first study to make them parallel with the other four:

> Jimmy/Timmy detected the blue box.
> Jimmy/Timmy saw the blue box.

These probes were given to 100 participants who passed the materials checks. The results are shown in Figure 3. With this slight revision, I now found that slightly more than half of the participants answered that "Jimmy saw the blue box"; this is a significant increase from the percentage selecting the corresponding answer in FAN's first study.[12]

[9] Neither difference was significant, although it was borderline significant for Timmy detected: Two-sample test for equality of proportions with continuity correction, χ^2=0.3798, df=1, p=0.2689, one-tailed (saw); χ^2=2.6841, df=1, p=0.05068, one-tailed (detected).

[10] The difference was not significant: Two-sample test for equality of proportions with continuity correction, χ^2=0.862, df=1, p=0.1766, one-tailed.

[11] One-sample test of proportion with continuity correction, χ^2=0.98, df=1, p=0.1611, one-tailed.

[12] Two-sample test for equality of proportions with continuity correction, χ^2=3.8632, df=1, p=0.02468, one-tailed.

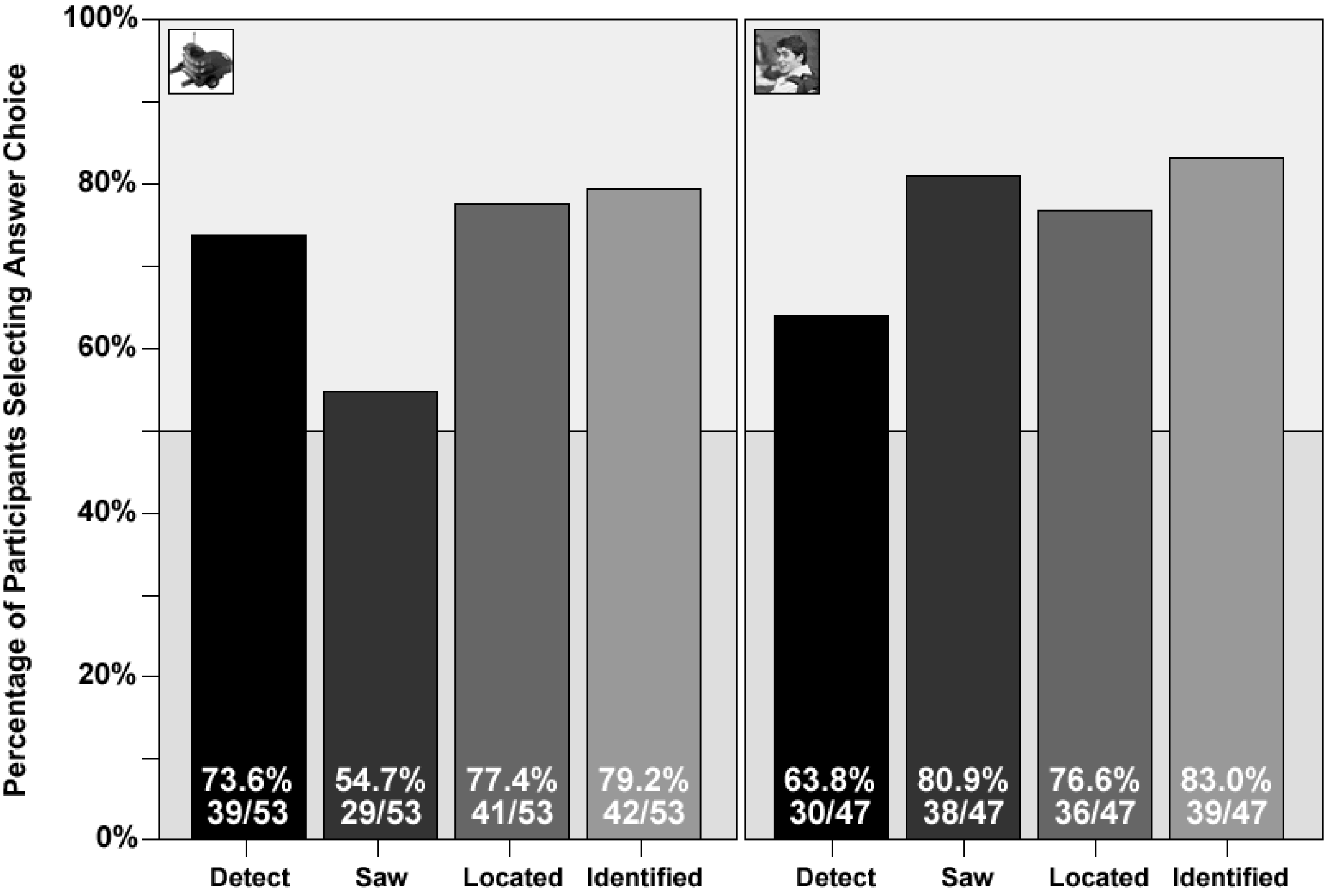

Figure 3. Results of Study 2: Revised statements in original ordering.

In my third study, I used the same probes as in Study 2, but changed the ordering of the answer choices. Participants were now given the six choices in the following fixed order:

Jimmy/Timmy saw the blue box.
Jimmy/Timmy located the blue box.
Jimmy/Timmy identified the blue box.
Jimmy/Timmy detected the blue box.
Jimmy/Timmy moved the blue box.
Jimmy/Timmy moved the red box.

These probes were given to 85 participants who passed the materials checks. The results are shown in Figure 4. Once again, a majority of the participants receiving the Jimmy probe selected

"Jimmy saw the blue box"; further, a larger percentage selected this answer choice than did in my second study. What's more, the percentage of participants selecting that answer choice was not significantly different from the percentage selecting that "Jimmy detected the blue box."[13]

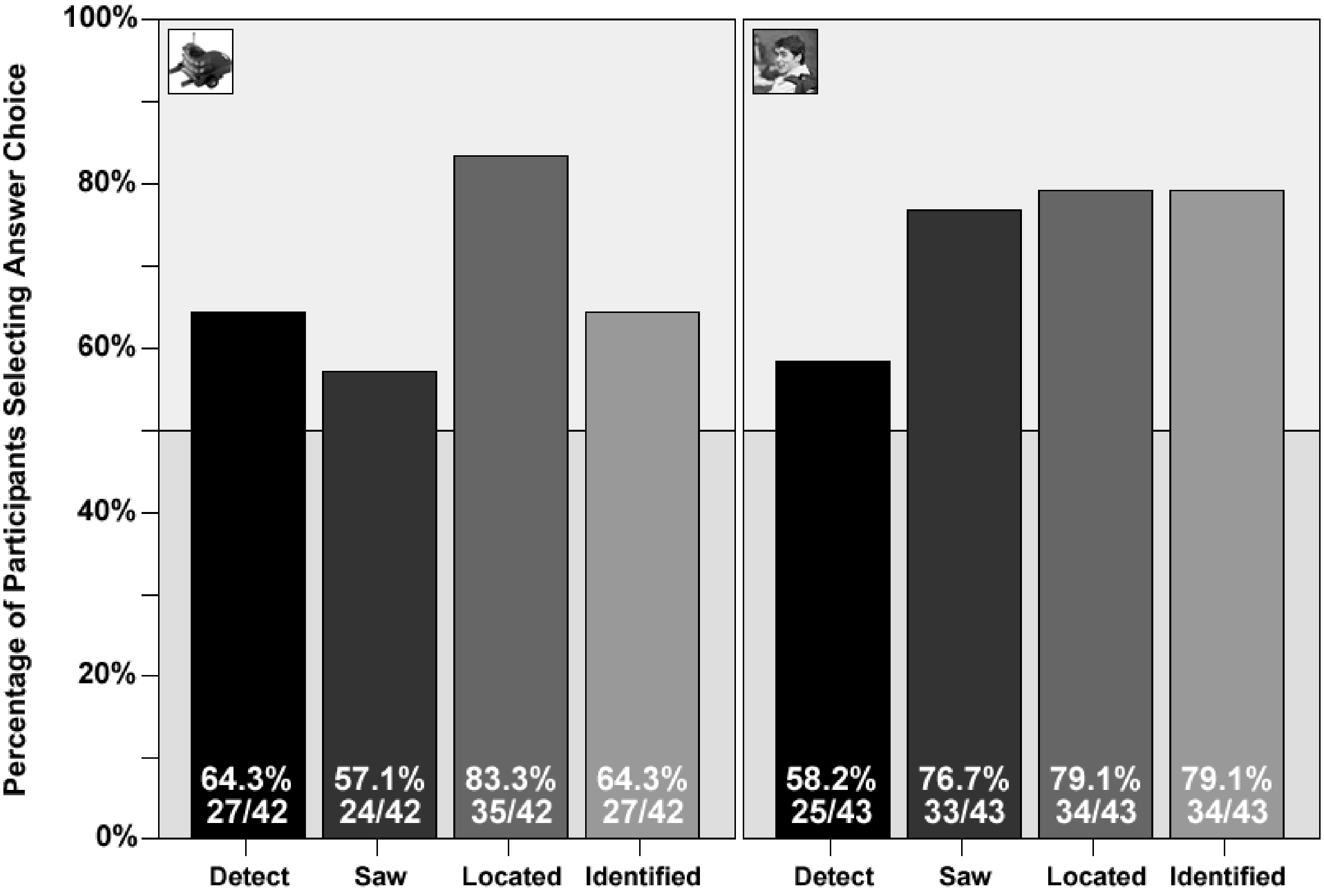

Figure 4. Results of Study 4: Revised statements in revised ordering.

As a final test, in my fourth study I followed a suggestion given by Shaun Nichols and removed the "detect" answer choice from the probes used in my previous study. These probes were given to 127 participants who passed the materials checks. The results are shown in Figure

[13] Two-sample test for equality of proportions with continuity correction, χ^2=0.1996, df=1, p=0.3275, one-tailed.

5. Yet again, a majority of the participants receiving the Jimmy probe selected "Jimmy saw the blue box," and this proportion was significantly above what one would expect by chance.[14] In fact, we see that a higher percentage of participants selected the saw answer choice than selected the detect answer choice.

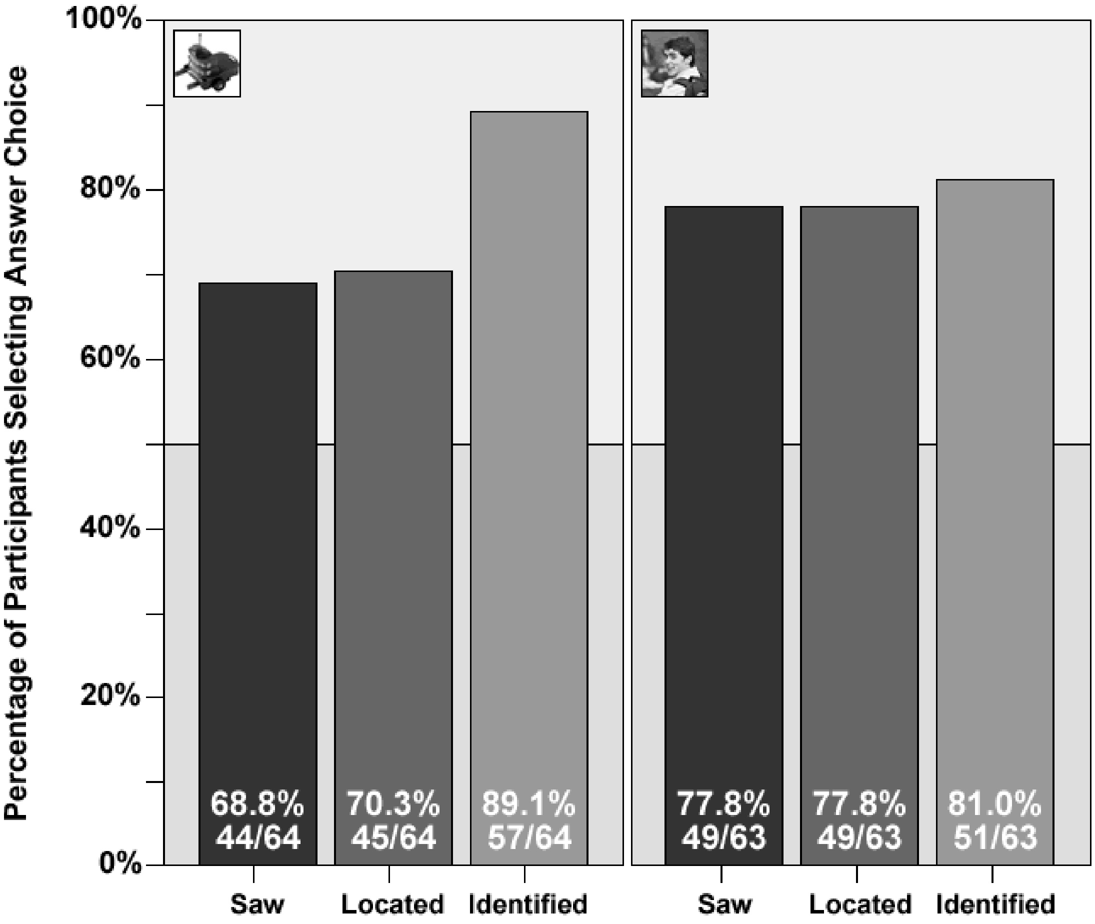

Figure 5. Results of Study 5: Revised statements without "detect" answer choice.

Across the five studies we have looked at—FAN's first study and the four follow-up studies reported in this section—we see a steady increase in the percentage of participants

[14] One-sample test of proportion with continuity correction, χ^2=8.2656, df=1, p=0.00202, one-tailed.

answering that Jimmy saw, as shown in Figure 6. This progression suggests that the worries about FAN's study that I raised in the previous section were well placed: As we increase the sample size and restrict to lay people, revise the answer choices to make the first two parallel with the rest, change the ordering of the answer choices, and finally remove the "detect" answer choice, we see a progressive increase in the percentage of participants answering that Jimmy saw. And at the end of this progression we find that people are significantly more likely to answer that Jimmy saw than not, even though they could otherwise indicate that Jimmy performed the color discrimination task by answering that Jimmy identified and/or located the relevant box. I take this to confirm the claim that lay people are generally willing to attribute mental states of seeing to the simple robot.

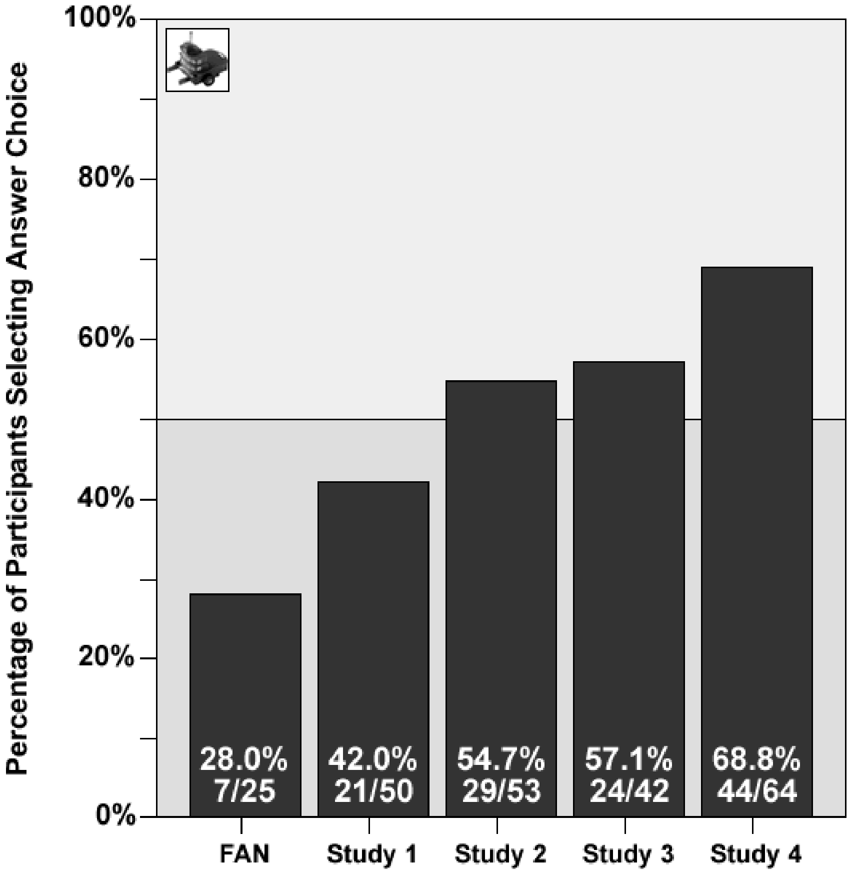

Figure 6. Results for "Jimmy saw" across five studies.

4. Let's Get Together

I began by noting that we can distinguish between positive and negative objectives in three prominent lines of investigation in the recent experimental philosophy of mind literature. Most importantly for the purposes of this chapter, we can distinguish between the negative hypothesis in the work of Sytsma and Machery (that the existence of phenomenally conscious mental states is not obvious from first-person experience with states like seeing red and feeling pain) and the positive hypothesis in the work of FAN (categorization of an entity as an AGENT via low-road processing produces a disposition to ascribe a wide range of mental states to that entity). These two hypotheses need not be seen as being at odds. In fact, I think that they can actually be seen as fitting together rather nicely: If FAN are correct, then low-road processing does not reflect the philosophical distinction between mental states that are phenomenally conscious and those that are not, which fits with the claim that this distinction is non-obvious. Put another way, the work of FAN can be taken to provide support for the claim that the existence of phenomenally conscious mental states is not obvious to low-road processing, while the work of Sytsma and Machery can be taken to provide support for the corresponding claim with regard to high-road processing.

Despite this, FAN argue in their contribution to this volume that we should expect high-road processing to generate a general disinclination to attribute a wide range of mental states to robots. And this claim is in conflict with the evidence provided by Sytsma and Machery. In response, FAN report on a new study that calls our results into doubt: They find that people are apparently less willing to ascribe mental states of seeing to the simple robot Jimmy than indicated by our previous studies. While I am a fan of FAN in general, and their agency model in particular, I do not think that these new results are compelling. Rather, I find that the structure of the questions asked by FAN were depressing attributions of seeing to both the simple robot Jimmy and the normal human Timmy. And the results of the four new studies that I reported in this chapter support that charge. I conclude that the evidence continues to support Sytsma and Machery's negative hypothesis, even as it continues to support FAN's agency model.

References

Arico, Adam, Brian Fiala, Rob Goldberg, and Shaun Nichols (2011). "The Folk Psychology of Consciousness." Mind & Language, 26(3): 327–352.

Atler, A. L., D. M. Oppenheimer, N. Epley, and R. N. Eyre (2007). "Overcoming intuition: Metacognitive difficulty activates analytic reasoning." Journal of Experimental Psychology: General, 136: 569–576.

Buckwalter, Wesley and Mark Phelan (forthcoming-a). "Phenomenal Consciousness Disembodied." In J. Sytsma (ed.), *Advances in Experimental Philosophy of Mind*. London: Continuum.

Buckwalter, Wesley and Mark Phelan (forthcoming-b). "Function and Feeling Machines: A Defense of the Philosophical Conception of Subjective Experience." *Philosophical Studies*.

Fiala, Brian, Adam Arico, and Shaun Nichols (2011). "On the Psychological Origins of Dualism: Dual-process Cognition and the Explanatory Gap." In E. Slingerland and M. Collard (eds.), Creating Consilience: Issues and Case Studies in the Integration of the Sciences and Humanities. Oxford University Press.

Fiala, Brian, Adam Arico, and Shaun Nichols (present volume). "I, Robot."

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

Gray, Heather, Kurt Gray, and Daniel Wegner (2007). "Dimensions of Mind Perception." *Science*, 315: 619.

Huebner, Bryce (2010). "Commonsense concepts of phenomenal consciousness: Does anyone *care* about functional zombies?" *Phenomenology and the Cognitive Sciences,* 9: 133–155.

Huebner, Bryce, Mike Bruno, and Hagop Sarkissian (2010). "What Does the Nation of China Think about Phenomenal States?" Review of Philosophy and Psychology, 1(2): 225–243.

Knobe, Joshua and Jesse Prinz (2008). "Intuitions about Consciousness: Experimental Studies." *Phenomenology and Cognitive Sciences*, 7: 67–85.

Reuter, Kevin (2011). "Distinguishing the Appearance from the Reality of Pain." Journal of Consciousness Studies, 18(9-10): 94–109.

Reuter, Kevin and Dustin Phillips, and Justin Sytsma (forthcoming). "Pain Hallucinations." In J. Sytsma (ed.), *Advances in Experimental Philosophy of Mind*. London: Continuum.

Sytsma, Justin (2009). “Phenomenological Obviousness and the New Science of Consciousness.” *Philosophy of Science*, 76(5): 958–969.

Sytsma, Justin (2010). “Dennett’s Theory of the Folk Theory of Consciousness.” *Journal of Consciousness Studies*, 17(3-4): 107–130.

Sytsma, Justin (forthcoming). “Revisiting the Valence Account.” *Philosophical Topics*.

Sytsma, Justin and Jonathan Livengood (2011). “A New Perspective Concerning Experiments on Semantic Intuitions.” *Australasian Journal of Philosophy*, 89(2): 315–332.

Sytsma, Justin and Jonathan Livengood (forthcoming). *The New Experimental Philosophy: An Introduction and Guide*. Broadview Press.

Sytsma, Justin and Edouard Machery (2009). “How to Study Folk Intuitions about Phenomenal Consciousness.” *Philosophical Psychology*, 22(1): 21–35.

Sytsma, Justin and Edouard Machery (2010). “Two Conceptions of Subjective Experience.” *Philosophical Studies*, 151(2): 299–327.

Sytsma, Justin and Edouard Machery (2012a). “The Two Sources of Moral Standing.” *Review of Philosophy and Psychology*, 3(3): 303–324.

Sytsma, Justin and Edouard Machery (2012b). “On the Relevance of Folk Intuitions: A Reply to Talbot.” *Consciousness and Cognition*, 21(2): 654–660.

Talbot, Brian (2012). “The Irrelevance of Folk Intuitions to the ‘Hard Problem’ of Consciousness.” *Consciousness and Cognition*, 21(2): 644–650.