# Folk Psychology, Consciousness, and Context Effects[*]

**Abstract**: Traditionally, the philosophical study of Folk Psychology has focused on how ordinary people (those without academic training in fields like Psychology, Cognitive Science, Philosophy of Mind, etc.) go about attributing mental states like *beliefs* and *desires*. Philosophers call such states *intentional states*. Recently, a body of work has emerged in the growing field of Experimental Philosophy that focuses on folk attributions of certain states not previously discussed in the Folk Psychological literature; namely the discussion is concerned with figuring out how (and whether) ordinary people go about attributing mental states of qualitative experience, or what philosophers might call *phenomenal consciousness*. The discussion has centered, largely, on two questions: Do the folk distinguish, in some sense, between states of phenomenal consciousness and nonphenomenal states (like intentional states)? And if they do distinguish between the two kinds of states, do the folk discriminate between different kinds of entities in their attributions of phenomenal states? This paper hopes to contribute to that discussion by discerning two primary hypotheses presently competing in the existing experimental philosophy literature, by presenting some experimental data that weigh on those hypotheses, and by offering a cognitive model of the processes underlying attributions of mental states.

## 1 Introduction

Within the Philosophy of Mind tradition, philosophers have often asked the question of how we, as humans, go about determining whether or not another entity has a mind. Historically, we can see explanations of how people ascribe minds put forth by such figures as St. Augustine (*De Magistro* (Augustine 1995)), John Stuart Mill (Mill 1865), and Thomas Reid (Reid1785), to name a few. Importantly, these figures were not merely interested in the ways we *ought* to go about determining whether an entity has a mind, nor were they concerned with the merely with analyzing the concept of 'mind' that is attributed (or withheld) in such judgments. Rather, they were all involved to some extent in a purely descriptive project aimed at explaining the actual processes by which humans, in fact, derive judgments of an entity's mindedness.

Today, philosophers call the everyday practice of attributing mentality 'Folk Psychology'. Since the last quarter of the twentieth century, researchers—philosophers and psychologists alike—have focused the Folk Psychology debate on the question of how we go about attributing particular mental

[*] Acknowledgements removed for blind review.

stats, such as beliefs and desires—which are generally called 'intentional states'. This discussion has largely been between Theory Theorists (Gopnick and Meltzoff 1997, Sellars 1956) and Simulation Theorists (Goldman 1989, Gordon 1986, Heal 1986),[1] who offer competing accounts of the mechanisms and processes underlying our everyday attributions of intentional states. This paper is not primarily a contribution to the topic of Folk Psychology so described.

More recently, philosophers, psychologists and cognitive scientists have begun to discuss a variation on the Folk Psychology project, which focuses on mental states that are not considered intentional states.[2] The question under consideration is a slight modification of that at issue in the debate between Theory theorists and Simulation theorists: whereas the debate has so far focused on folk attributions of mental states like beliefs and desires, the current debate focuses on folk attributions of mental states like 'feeling pain' or 'seeing read'. Such states are generally called 'phenomenal states'. This new Folk Psychology project asks 'how do we, as humans, go about attributing phenomenal consciousness?' This discussion is primarily concerned with two aspects of this question: do the folk distinguish, at some level, between intentional mental states and phenomenal mental states? and if they do, do the folk discriminate differently in their attributions of those mental states?

Because the new Folk Psychology project, and particularly its manifestation in experimental philosophy, is a relatively new philosophical discussion, there is relatively little literature on the topic. That said, we can nonetheless discern two primary hypotheses that fall out of the existing body of work: one hypothesis states that our commonsense psychology includes a specific restriction against attributing to groups states of phenomenal consciousness; the other hypothesis claims that our commonsense psychology is sensitive to context effects but is not specifically biased against groups having phenomenal consciousness. These hypotheses, fortunately, are entirely testable by experimental methods. Each makes its own empirical predictions, and the two sets of predictions differ in a way that can be straight-forwardly tested by simply surveying commonsense intuitions. In this paper I attempt to lay out the two

[1]Though there is also a small minority of theorists who subscribe to a Phenomenological Theory of Mind (Gallagher 2005, Zahavi 2005).
[2] See Gray, Gray, and Wegner (2007), Robbins and Jack (2006), for instance.

hypotheses in more detail, including their implications and predictions, after which I will introduce new empirical data that directly bear on those hypotheses. Finally, I'll present and discuss a cognitive model (the 'Agency Model'), which I take to be the best account of the process underlying attributions of mental states, and which explains mental state attribution in a way that best accommodates the empirical data.

## 2 Existing Empirical Literature

I'd like to briefly discuss three research projects carried by experimental philosophers that focus on the Folk Psychology of phenomenal consciousness: Joshua Knobe and Jesse Prinz's "Intuitions About Consciousness" (2008), Bryce Huebner, Michael Bruno, & Hagop Sarkissian's "What Does the Nation of China Think about Phenomenal States?" (ms), and Justin Sytsma and Edouard Machery's "How to Study Folk Intuitions About Phenomenal Consciousness" (forthcoming).

Knobe & Prinz argue from a series of experiments for the conclusion that the folk have a built-in restriction that prevents them from attributing phenomenal consciousness to groups. There is insufficient room here to present all of the studies Knobe & Prinz offer in support of their thesis, but a quick discussion of two will suffice to capture their argument. In one study, Knobe & Prinz present subjects with two kinds of sentences: one attributing phenomenal states to groups, and one attributing nonphenomenal states to groups. For instance:

Phenomenal:
*Acme Corp. is getting depressed.*
*Acme Corp. is feeling excruciating pain.*
*Acme Corp. is feeling regret.*
Nonphenomenal:
*Acme Corp. has just decided to adopt a new marketing plan.*
*Acme Corp. intends to release a new product this January.*
*Acme Corp. regrets its recent decision.*

They then asked subjects to rate each sentence as sounding 'weird' or sounding 'natural'.[3] Subjects found phenomenal attributions to groups significantly 'weirder' than nonphenomenal attributions. The second study then tested whether this bias against attributing to groups phenomenal consciousness was driven by

[3] Responses were recorded using a 7-point Likert scale (1='sounds weird', 7='sounds natural').

groups being judged to be inappropriate loci of, say, being upset or by groups being judged to be inappropriate loci of phenomenal experience (i.e., of *feeling* upset). Knobe & Prinz once again gave subjects two kinds of sentences: one attributing the state of 'upsetness' (or 'regret'), and one attributing the state of 'feeling upset' (or 'feeling regret'); again, subjects were asked to rate sentences for sounding 'weird' or 'natural'. The average rating for sentences that included the 'feeling' locution was significantly lower than the average rating for sentences without the 'feeling' locution, indicating that the driving force behind the original asymmetry is a bias against granting groups the capacity to *feel* or *experience* various mental states. Knobe & Prinz conclude: In other words, the folk have a specific restriction on ascribing phenomenal consciousness to groups (as Knobe & Prinz put it, "a restriction against agents that are composed of other agents." (p. 76)).

Huebner, Bruno, & Sarkissian later replicated the Knobe & Prinz results, finding that folk, both from the West and from East Asia, treat groups and individuals differently when it comes to attributing phenomenal consciousness. Using the same methodology as Knobe & Prinz's first study, Huebner et al. asked subjects (some, Westerners at the University of North Carolina—Chapel Hill; some, East Asians at Hong Kong University) to rate different kinds of sentences as sounding 'weird' or 'natural'. Like Knobe & Prinz, Huebner & colleagues presented their subjects with some sentences attributing phenomenal consciousness to groups and with some sentences attributing nonphenomenal states to groups. Interestingly, they also presented subjects with some sentences attributing phenomenal consciousness to individuals, as well as some sentences attributing nonphenomenal states to individuals:

Group Phenomenal:
*The Ming Dynasty felt relief after the rebellion was quelled.*
Group Nonphenomenal:
*The Ming Dynasty thought that China was the greatest country in the world.*
Individual Phenomenal:
*Tanya feels relieved after her mother's successful surgery.*
Individual Nonphenomenal:
*Tanya believes that China is the greatest country in the world.*

While they found that their subjects also rated phenomenal attributions to groups significantly lower than phenomenal attributions to individuals, the emphasis of their study was that this effect is significantly greater in Westerners than in East Asians:

> [While] our American volunteers found ascriptions of mental states to *individuals* considerably more acceptable than ascriptions of mental states to *collectivities*, volunteers from Hong Kong did *not* see such a considerable [difference] between the two. (emphasis theirs) (8)

Although Huebner et al. found a difference in the degree to which East Asians and Westerners prefer attributions of phenomenal states to individuals over phenomenal attributions to groups, it is important to keep in mind that East Asians nonetheless rated phenomenal attributions to groups significantly lower than both phenomenal attributions to individuals and nonphenomenal attributions to groups. In short, their results mirror those found in the Knobe & Prinz studies.

Justin Sytsma & Edouard Machery, on the other hand, argue from their own series of experiments that the Knobe & Prinz data are ambiguous, and that their results can be explained by something other than a specific restriction against attributing phenomenal consciousness to groups. Sytsma & Machery have two main lines of argument, though in the name of discursive parsimony, I will discuss only one.[4] In one section, they argue that since the stimuli utilized in the Knobe & Prinz experiments did not form what are referred to as 'minimal pairs' (see discussion below), there are multiple possible explanations for the asymmetry in responses. To demonstrate this point, they ran an experiment in which they presented subjects with four sentences:

> *Acme Corporation is feeling upset about the court's recent ruling*
> *Acme Corporation is upset.*
> *Acme Corporation is feeling regret about its recent decision.*
> *Acme Corporation regrets.*[5]

Subjects rated these sentences on a 7-point Likert scale, from "clearly weird" to "clearly natural". Contrary to the Knobe & Prinz results, subjects rated the phenomenal attributions higher than

[4] Sytsma & Machery's second argumentative line is that Knobe & Prinz may have actually been measuring the folk recognition of functional and behavioral differences between groups and individuals. In support of this hypothesis, they present data from their own experiments, the details of which I will not go into here.

[5] Notice that these sentences essentially invert the structure of the sentences used in Knobe & Prinz's second study.

nonphenomenal attributions (though the difference was not significant for 'regret' and only marginally significant for 'upset'). According to Sytsma & Machery, this data undermines the idea that Knobe & Prinz were tracking a cognitive bias specifically against attributing phenomenal consciousness to groups, and suggests that the original data consist with an alternative explanation:

> These results support our hypothesis that in Knobe and Prinz's study 4, subjects' preference for sentences formed with "feeling upset" and "feeling regret" over, respectively, sentences formed simply with "is upset" and "regret" results from the fact that in the former sentences, but not in the latter sentences, the verb was followed by a prepositional phrase. (13)

Their point is that much of the existing data can be better explained by something other than a built-in restriction against certain sorts of attributions.

## 3 Methodological Worries

As interesting and important as all of these studies are, each of them is missing one of two methodological features that I take to be of crucial importance to the empirical study of the Problem of Other Conscious Minds. First, in order to draw any conclusions about judgments of group consciousness, we have to collect data that include judgments on more than just groups. Given that we are attempting to justify the conclusion that people dislike attributions of phenomenal consciousness *in virtue of those attributions involving groups*, we have to establish that people respond differently when those attributions involve non-group agents (or individuals). Thus, in order to draw any group-specific conclusions from our experiment, it must be designed such that data are collected on attributions of consciousness both to groups and to individuals—and, of course, this data must show either a significant difference or a significant correlation between the two sets of responses if we are to arrive at any positive conclusions. Let's call this the "crucial comparison."[6]

[6] The point might be put more straight-forwardly that every experimental variable requires a control condition. For studies aimed at testing the impact of phenomenal consciousness (variable 1) being predicated of groups (variable 2), we need two controls: one control to compare with phenomenal consciousness (e.g., nonphenomenal states) and one control to compare with groups (e.g., individuals).

Second, we must be careful to utilize "minimal pairs" in our stimulus materials.[7] Allowing differences between stimuli beyond what is necessary to test the dependent variable opens the data to multiple possible interpretations, since we could not be certain that any effect was due to the independent variable rather than to the other differences between stimuli. Thus, in order to draw any definite conclusions about the cognitive process behind attributions of consciousness, we must be careful to maximize the similarity between stimuli.

Knobe & Prinz, while being primarily responsible for initiating the experimental philosophical contribution to the Problem of Other Conscious Minds, did not use minimal pairs and did not include what I'm calling the crucial comparison. Huebner, Bruno, & Sarkissian did include the crucial comparison, but they did not use strictly minimal pairings. Sytsma & Machery, by contrast, did use minimal pairs but did not include the crucial comparison (which makes their appeal to functional and behavioral differences between individuals and groups somewhat mysterious. How would such a difference explain the asymmetry between phenomenal attributions *to groups* and nonphenomenal attributions *to groups* without begging the question that phenomenal consciousness is a uniquely individual function/behavior?). As such, these two methodological worries were critically-formative in designing my own attempt to answer the question of whether or not the folk consider it appropriate to attribute phenomenal consciousness to groups. The next section lays out the details of that experiment and presents the results thereof.

## 4 New Experimental Data

The question under consideration is how the cognitive process underlying attributions of phenomenal consciousness respond to groups. As such, we have to compare judgments of attributions of *phenomenal consciousness* to judgments of attributions of *nonphenomenal states*.[8] And, as noted above, insofar as we

[7] Minimal pairs are stimuli that are maximally similar save for the one variable (i.e., the 'independent variable') that is being tested for effect.

[8] The accepted wisdom in the literature to date is that sentences are phenomenal attributions insofar as they use phenomenal language, namely when they include locutions like 'is feeling' or 'is experiencing' when describing the

want to draw conclusions about processing attributions of phenomenal consciousness to *groups*, we have to include the crucial comparison class of attributions to *individuals*.[9] There are, then, at least two dimensions along which we are testing folk intuitions: attributions to groups vs. attributions to individuals; and attributions of *phenomenal consciousness* vs. attributions of *nonphenomenal states*. I added a third dimension, which requires a brief explanation.

One reasonable response to the Knobe & Prinz studies, similar to that proposed by Sytsma & Machery, is that the effect they found could be chalked up almost entirely to uncontrolled-for features of their stimuli, to their failure to utilize minimal pairs. One version of such a response interprets the asymmetry in subjects' responses as an artifact of, specifically, the asymmetrical presence (or absence) of some context-providing information in their stimuli.[10] The sentences that Knobe & Prinz asked subjects to rate were not minimal pairs, so we cannot be certain that the differences between responses was an effect caused by the attribution involving phenomenal consciousness, rather than an artifact of the other sentential differences. However, the study by Huebner, et al., on the other hand, also gives some empirical support for thinking that there really is some cognitive bias against attributing to groups certain psychological states (namely, phenomenal consciousness) and that this bias is driving the asymmetrical data. We have, then, two competing hypotheses:

> *The Context-Sensitive Hypothesis*: People's willingness to attribute states of phenomenal consciousness is sensitive to the presence (or absence) of contextual information but is not specifically biased against attributing to groups states of phenomenal consciousness.
>
> *The Biased Hypothesis*: People are reluctant to attribute states of phenomenal consciousness to groups because they have a group-specific cognitive restriction against attributing phenomenal consciousness. [11]

---

psychological state. I follow in this tradition, though the opening paragraphs of the discussion section will offer some reason for thinking this assumption is unwarranted.

[9] This crucial comparison will also need to include attributions of *nonphenomenal states* to individuals.

[10] The attribution sentences that Knobe & Prinz presented to their subjects were non-minimal pairs because their nonphenomenal attributions included additional information (about the context of attribution) that their phenomenal attributions lacked:
*'Acme Corp is upset about the court's recent ruling'* vs. *'Acme Corp is feeling upset.'*

[11] Clearly, these two theses need not be entirely incompatible (see Discussion).

Fortunately, both of these hypotheses make clear empirical predictions. The Context-Sensitive hypothesis predicts that (a) once we balance attributions of both phenomenal and nonphenomenal states such that they include contextual information, we should see the asymmetry between phenomenal and nonphenomenal mostly, if not entirely, vanish; and (b) that people will rate attributions of phenomenal consciousness to groups without contextual information significantly lower than attributions of phenomenal consciousness to groups that include contextual information. The Biased hypothesis, on the other hand, predicts that, even if we include contextual information within the attributions of phenomenal consciousness to groups, people should nonetheless rate them (a) significantly lower than attributions of nonphenomenal states to groups and (b) significantly lower than attributions of phenomenal consciousness to individuals. Thus, this experiment tested intuitions along a third dimension: attributions *with* context vs. attributions *without* context.

**Method**

*Participants*

Participants were 61 University of Arizona undergraduate students enrolled in an introductory philosophy course. Participation was entirely voluntary.

*Materials*

Two surveys were prepared for the study. One survey included 12 attributions of mental states to groups; the other survey included 12 attributions of mental states to individuals. Sentences were counterbalanced to accommodate for order-effects. The attributions for both surveys broke down into four categories as follows:

- 3 attributions WITH the 'feeling' locution, but WITHOUT contextual information:
  e.g., "McDonald's is feeling upset."
  "Donald is feeling upset."
- 3 attributions WITH the 'feeling' locution and WITH contextual information:
  e.g., "The Giants felt proud after winning the Super Bowl."
  "Manning felt proud after winning the Super Bowl."
- 3 attributions WITHOUT the 'feeling' locution and WITHOUT contextual information:

e.g., "The Board of Trustees was relieved."
"Thomas was relieved."

- 3 attributions WITHOUT the 'feeling' locution but WITH contextual information:
  e.g., "The consulting firm is depressed about being dumped."
  "Michael is depressed about being dumped."

*Procedure*

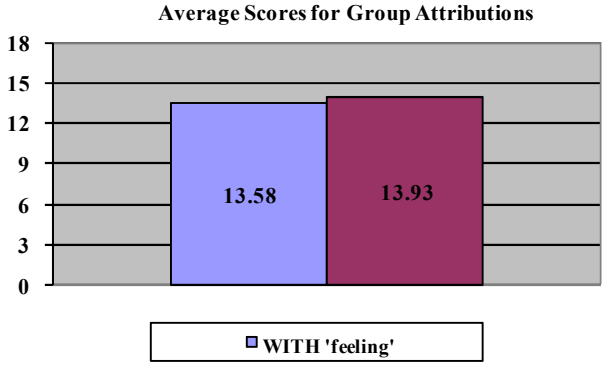
Subjects were separated into two conditions: the Group condition, which received the survey attributing mental states to group agents; and the Individual condition, which received the survey attributing mental states to individuals. They were then asked to rate each sentence on a 7-point Likert scale (1 = 'sounds weird', 7 = 'sounds natural'[12]). Response scores for each category were then calculated by summing the responses of all three attributions, such that each participant produced four scores (one for each category of attribution). We could thus do a between-subjects comparison across those four dimensions to see (a) whether there is a difference between judgments of attributions to groups and judgments of attributions to individuals, and (b) if such a difference were observed, whether the source of that difference could be traced back to the presence of contextual information or to the attribution including phenomenal language (i.e., with the 'feeling' locution).

*Results*

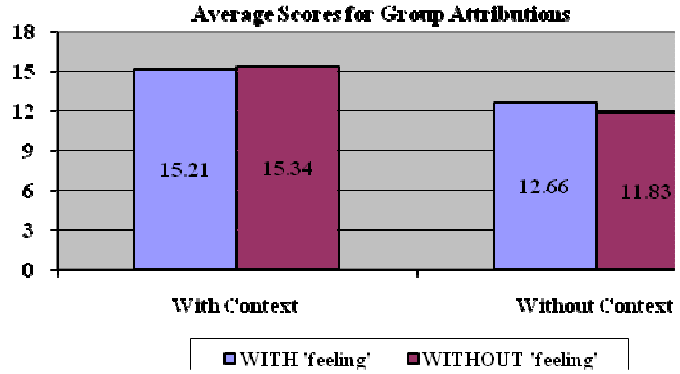
Contrary to what the Biased hypothesis predicted, there were no differences between group attributions WITH 'feeling' and group attributions WITHOUT 'feeling'.[13]

[12] This question has become ubiquitous in experimental philosophy since Joshua Knobe's initially using it in his now-famous Help/Harm cases. See Knobe 2003.

[13] $t(114) = .458$, $p = .648$

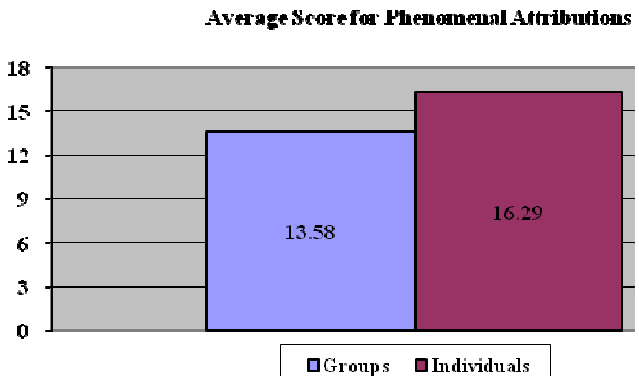



And as the Context-Sensitive hypothesis predicted, when both (or neither) attributions include a context, the difference between folk judgments on attributions WITH 'feeling' and attributions WITHOUT 'feeling' vanishes.[14]

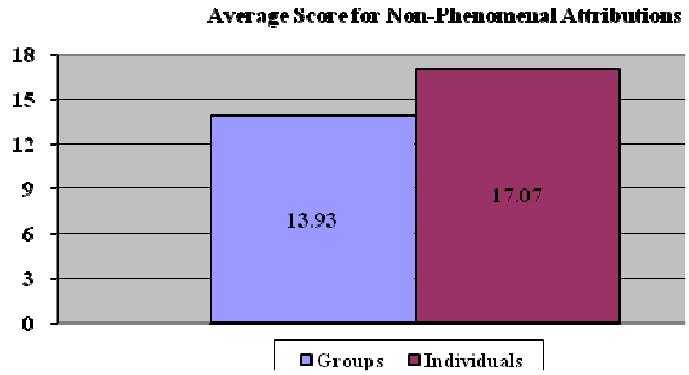



I did find, however, that group attributions in both conditions were rated significantly lower than the corresponding attribution to individuals. Prediction (b) of the Biased hypothesis—that even when context is controlled for, subjects will rate group attributions WITH 'feeling' lower than individual attributions WITH 'feeling'—turns out to be correct. That is, subjects rated attributions to groups that included phenomenal language significantly lower than they rated attributions to individuals that included phenomenal language.[15]

[14] Without Context: $t(56) = -.136$, $p = .893$. With Context: $t(56) = .856$, $p = .396$.
[15] $t(120) = 3.98$, $p<.001$



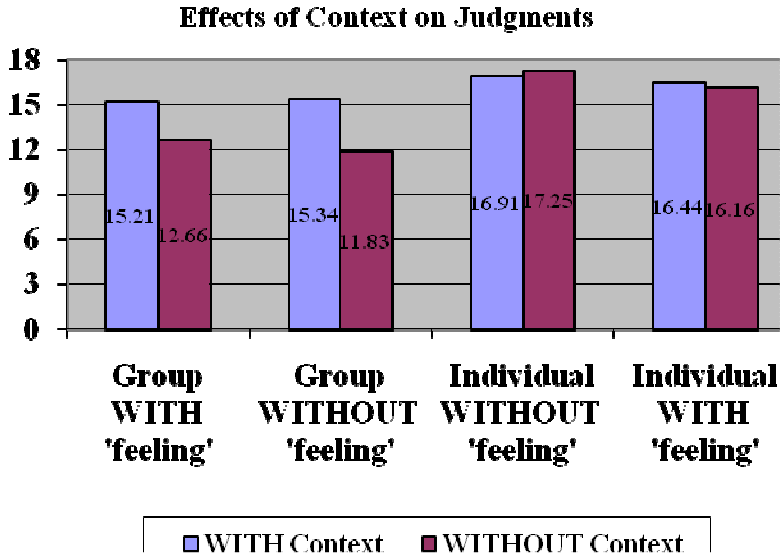

However, contrary to the Biased hypothesis, this asymmetry does not appear to be about phenomenal consciousness *per se*. Rather, it seems that the folk have a more general bias against group psychology that includes, but is not limited to, group attributions including phenomenal language. So much is evidenced by the comparison between judgments on group attributions WITHOUT 'feeling' and individual attributions WITHOUT 'feeling'. Just as the folk rate group attributions with phenomenal language lower than those same attributions to individuals, they also rate group attributions without phenomenal language significantly lower than corresponding attributions to individuals.[16]



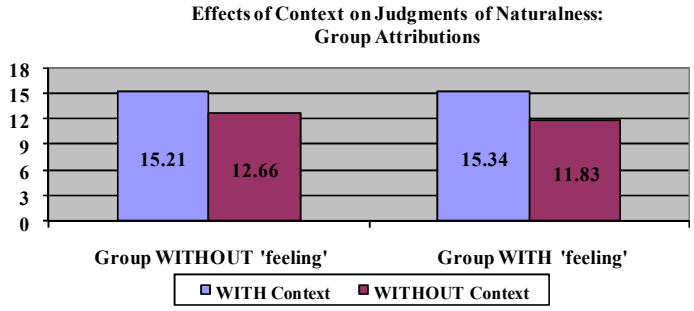

Perhaps the most interesting result of the experiment, however, was the effect that including or excluding the additional contextual information had on subjects' ratings. The chart below illustrates what we might call "the context effect". Namely, the data reveal just how much *more* sensitive judgments of sentences

[16] $t(120) = 4.73$, $p<.001$

are to contextual information when psychological states are attributed to groups than when psychological states are attributed to individuals.

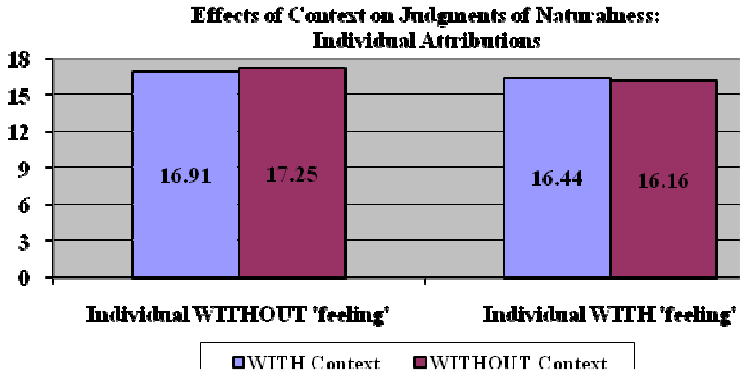



When we look at the difference between sentences that included contexts and sentences that did not, we see a highly significant difference, but only when the attribution is to a group. When sentences attributing mental states to groups included contextual information, ratings were significantly higher than when they did not include context.[17]




[17] $t(114) = 4.35$, $p < .001$, for all attributions; $t(56) = 2.525$, $p = .014$, for phenomenal; $t(56) = 3.61$, $p < .001$, for nonphenomenal.

Attributions to individuals, on the other hand, showed no difference between 'with context' and 'without context' conditions.[18]




*Discussion*

There are, I think, three features of these data that merit some in-depth consideration: first, there is the fact that subjects treated group attributions of phenomenal consciousness and group attributions of nonphenomenal states the same; second, the difference between ratings of attributions WITH context and ratings of attributions WITHOUT context was only significant when those attributions involved groups; and third, although group attributions WITH context were rated significantly more natural than those WITHOUT context, they were nonetheless still rated as significantly less natural than corresponding attributions to individuals.

Recall that the Biased theorist interprets data from previous studies as showing that the folk distinguish (even if only tacitly) between phenomenal consciousness and other, nonphenomenal kinds of mental states.[19] They draw this conclusion from the previously observed asymmetry in folk judgments of sentences attributing the two kinds of consciousness: different judgments, the line of reasoning goes, suggest the recognition (on some level) of two distinct types of consciousness in the mental state

[18] $t(126) = .052$, $p = .959$, for all attributions; $t(62) = .411$, $p = .682$, for phenomenal; $t(62) = .326$, $p = .745$, for nonphenomenal.

[19] Put another way, they take their data to support the notion that the folk have a distinct concept of phenomenal consciousness, which is withheld (for some reason or another) from group agents. Knobe & Prinz, for instance, argue that the folk distinguish between the psychological state of 'upsetness' and the phenomenal experience of 'feeling upset.'

attributions; there is, in fact, a clear difference in judgments of attributions that tracks the (philosopher's) distinction between phenomenal and nonphenomenal; thus, on some level, the folk distinguish between phenomenal consciousness and nonphenomenal states. This line of reasoning, of course, is challenged by the current observation that, once the stimuli have been balanced to form minimal pairs, the folk *do not* rate group attributions of phenomenal consciousness any differently than group attributions of nonphenomenal states. Of course, we can interpret this particular finding in one of two ways; however, I suspect either one will prove problematic for the Biased theorist.

On one interpretation of the fact that my subjects rated phenomenal and nonphenomenal attributions nearly identically, the data merely show that adding the locution "is feeling" (or "is experiencing") to an attribution does not suffice to make that attribution a *phenomenal* attribution. On the other interpretation, the data seem to show that the folk do not distinguish between the two kinds of metal states. If the former interpretation is correct and attributions with the "is feeling" locution are not in fact phenomenal attributions, then the conclusion drawn from the Knobe & Prinz and the Huebner et al. data is not in fact supported by that data. We cannot conclude from the observed asymmetry in judgments *anything* about whether the folk distinguish between phenomenal and nonphenomenal states, since none of the sentences presented to them were phenomenal attributions. On this interpretation, experimenters (myself included) have yet to adequately present subjects with phenomenal attributions. On the other hand, even if adding the "is feeling" locution *does* suffice to make an attribution phenomenal in nature, the Biased conclusion is still undersupported by the data. That is, if attributions that include the "is feeling" locution count as phenomenal attributions, then the current data show that, once the sentences have been balanced to form minimal pairs, the folk do not treat phenomenal attributions any differently than nonphenomenal attributions. On either interpretation, the Biased theorist run squarely into trouble.

Having discussed the first point of interest, I would like to briefly digress and say a quick word or two about cognitive processes, before moving on to the second point of interest in our data. The digression focuses on the role that these studies play in our understanding of the human cognitive system. Although none of the literature on the Problem of Other Conscious Minds discussed so far includes any cognitive

model (or a positive proposal for solving said problem),[20] Knobe & Prinz briefly discuss what seems to be a constraint on any such model. In one study Knobe & Prinz ask subjects to explain either why a person would utilize ascriptions of memory or why a person would utilize ascriptions of consciousness. In the first (memory) condition, subjects responded with typical folk psychological explanations: "the man would be interested in ascribing a capacity for memory because this ascription could enable him to predict, explain or control behavior."(82) In the second (consciousness) condition, though, subjects responded without any reference to prediction, explaining or controlling behavior; rather, every subject responded along the lines of moral considerations. From this (and their previous data), Knobe & Prinz conclude that the process underlying folk judgments of consciousness attributions should not be understood as a whole, but should be broken down into two parts: one part attributing the state, and one part attributing the phenomenal consciousness of that state.

Although they don't talk about it in these terms, Knobe & Prinz's account suggests that any satisfactory cognitive model needs to include a separate and/or additional cognitive component in the process underlying attributions of other conscious minds (compared, that is, to the process underlying the attribution of other minds *simpliciter*). While the attribution of nonphenomenal states, on their account, is underwritten by traditional folk psychological considerations/processes, attributions of phenomenal states require an additional process of moral consideration.

This story parallels the one told in recent work by Philip Robbins and Anthony Jack (2006), who build on Daniel Dennett's notion of an *intentional stance* to develop what they call a *phenomenal stance*. Whereas Dennett argues that attributing an object intentional states (i.e., beliefs and desires) falls out of a need to explain or predict its behavior (when teleological and physical stances are too difficult or time-consuming), Robbins & Jack argue that attributing an object the ability to have phenomenal experiences

[20]Bryce Huebner (forthcoming) suggests a cognitive model insofar as he argues that commonsense (folk) judgments of nonphenomenal attributions are underwritten by a functionalist concept of 'consciousness' while judgments of phenomenal attributions and emotional attributions involve entirely distinct concepts of 'consciousness'. It's not clear to me, however, just how Huebner's account actually explains the process by which we determine whether an entity is capable of phenomenal experience. After all, on what grounds do we determine whether to utilize the functionalist concept of 'consciousness' rather than others?

falls out of empathizing with it and seeing it as "a potential target of moral concern." (70) Like Knobe & Prinz, then, Robbins & Jack think that attributing phenomenal consciousness is intimately tied to moral considerations; however, unlike Knobe & Prinz, Robbins & Jack take the cognitive capacity associated with phenomenal attributions to be *entirely* distinct from that associated with nonphenomenal attributions:

We noted above that the intentional stance is semantically, functionally, and neurally distinct from the physical stance. The phenomenal stance is distinct from its intentional and physical counterparts in all of these ways. (74)

Even though Robbins & Jack differ (perhaps) from Knobe & Prinz in the degree to which they think the processes overlap, the two accounts agree that judgments concerning phenomenal attributions require an additional cognitive component that is intrinsically connected with moral concerns.

The current study, however, presents some reason to doubt that there exists a fundamentally distinct process or additional component underlying attributions of phenomenal consciousness. Nothing in the present study suggests anything about additional or alternative processes behind judgments of phenomenal ascriptions compared to the process subserving judgments of nonphenomenal ascriptions. Put even more strongly, one might think that the current results show that whatever processes underlie folk attributions of phenomenal consciousness are largely similar to those underlying attributions of nonphenomenal states. Recall that the pattern of responses for phenomenal attributions was largely identical to the pattern of responses for nonphenomenal attributions, across the other conditions (Individual vs. Group, WITH Context vs. WITHOUT Context). I take the close correlation, in the present study, between responses on attributions of phenomenal consciousness to responses on attributions of nonphenomenal states to suggest that such judgments are subserved by either a single cognitive process (or single set of processes) or two computationally-similar cognitive processes (or sets of processes). If they are not the numerically same (set of) processes, then at the least they seem to be something like parallel (sets of) processes. Admittedly, this is not an outright refutation of the Knobe & Prinz/Robbins

& Jack story, but the current data does seem to once again place the burden of proof back on the shoulders of my theoretical opponents.

The above discussion of cognitive processes brings us to the second point of interest: the fact that only ratings of group attributions were affected by the addition of context. This second point is interesting because it reveals a functional component of the cognitive process underlying attributions of psychological states that has received little, if any, attention in the recent literature. Admittedly, framing effects are nothing new to either psychologists or philosophers, and if all that I was trying to show was that attributions of consciousness are subject to such effects, there would be little of interest here. However, what makes this data noteworthy and what calls for some sort of explanation is that only certain entities, namely groups, seem to be susceptible to the effect. The interesting point, the point that requires some sort of explanation, is that attributions of psychological states to individuals are so robust, so immune to context effects, while attributions of psychological states to groups are so fragile and easily effected by context. In what follows, I will present a model of human cognition that I think does a good job of explaining the data from this study, as well as data from other studies. It's called the Agency Model.

## 5 The Agency Model

As has been argued elsewhere (Arico, Fiala, Goldberg, & Nichols, *in prep*), there is reason to think that people are predisposed to attribute consciousness to other things largely because they have categorized the thing as an AGENT, and that whether or not a thing is categorized as an AGENT depends largely (though not exclusively) on whether the entity displays certain surface features. For instance, if the entity has eyes, displays certain motion trajectories, or interacts contingently with another individual, then that entity will activate in an observer the pre-potent disposition to categorize the entity as an AGENT; and when something is categorized as an AGENT, it is categorized as something that tends to have conscious experiences.[21] That is all to say that evolution seems to have designed the human cognitive system such

[21] This model is built largely on the model developed by Susan Johnson and colleagues (Johnson 2003, Johnson, Shimizu, and Ok 2007), which draws on work by Amanda Woodward (1998). Johnson hypothesizes that these

that things with certain surface features dispose us towards attributing to them, among other things, phenomenal consciousness. Importantly, included in this constellation of "other things" are goals and intentions, both of which are typically understood as nonphenomenal psychological states. Thus, the Agency Model presents a picture of cognition in which phenomenal and nonphenomenal attributions are generated by one and the same process. What remains, at this point, is to explain (i) how, on the Agency Model, we can make sense of the bias against attributing psychological states (in general) to groups, and (ii) why this bias is largely assuaged by additional contextual information. I will address each in turn below.

In general, groups do not display any of the typical AGENT features: while the component individuals in a group certainly possess eyes, move in certain trajectories, and contingently interact, the group *qua* collective entity rarely appears to have a pair of eyes, or to move in a distinct trajectory, or to contingently interact (especially those groups that we rarely observe directly at all, such as the more abstract variety: nations, corporations, bureaus, departments). And in the absence of these features, the cognitive mechanism is such that we are not typically disposed to attribute phenomenal consciousness. Thus, in general, groups are not the sorts of things that dispose us to attribute the ability to consciously experience. However, on those rare occasions when groups *do* appear to posses one of these surface characteristics, we see a willingness on the part of observers to attribute mentality. Paul Bloom and Csaba Veres (1999), for instance, showed subjects animations in which clusters of shapes moved in specific motion trajectories, which were similar to the motions of individual shapes in the original Heider-Simmel (1944) illusion. They found that subjects in their experiment used the same kind of intentional language to describe the motions of clusters of shapes as is often used to describe the motions of individual shapes (e.g. "'The blue dots would not let the green rectangles pass. However the green rectangles did not seem to mind and didn't try that hard. […] The blue dots seemed frustrated by this and still tried to get to the green rectangles." (Bloom & Veres, B8)). So, when groups are small enough and arranged so as to be

---

surface cues dispose infants to classify an entity as an AGENT, which then disposes them to display a variety of behaviors, including gaze-following, imitation, and the attribution of goals and goal-directed behavior. The Agency model essentially just adds to this constellation of behaviors the attribution of consciousness.

seen to move uniformly and in specific ways, they seem to trigger our AGENT detection process. Certain kinds of groups, then—such as bands, sports teams, classes, etc.—might be more easily thought of as possible loci of conscious experience than large groups like corporations and countries would be.[22]

Now, most of our everyday experiences of the kind of surface features mentioned above involve other individuals, so it is fairly natural to think of individuals as AGENTS. It is less often that we encounter those cues as involving large group agents (where the group is understood distinctly from the numerous individuals composing it), so it is less natural to think of groups as AGENTS. As instinctual attributions of consciousness go, then, we should expect just the kind of disparity between groups and individuals that we observe experimentally: we simply aren't conditioned to categorize groups as AGENTS, and thus aren't conditioned to attribute to them consciousness.

Fortunately, surface cues aren't the only means of categorizing a thing as an AGENT (although they are, I suspect, the most common). The path to AGENT categorization might run through more inferential territories in higher cognition. For instance, we might be told, through testimony from a typically-reliable source, that entity *X* has hopes, dreams, and desires and can think thoughts and feel pain. Because the testimony is from a reliable source, we come to believe that *X* is conscious despite the fact that we have never observed *X*'s surface features. Also, we might see someone interacting with a thing (say an object that is beyond our view) in a manner that is typical of interaction with an AGENT. Again, without the thing itself having triggered our AGENT-detector, we have nonetheless inferred from another Agent's observed behavior that the thing has consciousness. Likely there are other "side door" ways to categorizing a thing as an AGENT, but the point of emphasis here is that there are alternative methods of categorizing an entity as an AGENT that appeal to higher cognition and inferential reasoning processes.

All of this relates back to our question about how additional context impacts judgments of consciousness attributions, at long last, in that we might understand the additional, contextual information in sentences

[22] Of course, this is an empirically verifiable claim, though to my knowledge no work has been done on it specifically. In the Huebner, et al. paper, one stimulus included a group of just three members ('Destiny's Child is feeling insecure after its poor performance last night."), but obviously we shouldn't put too much weight on one data point.

attributing mental states to groups as serving as one of these "side doors" to categorizing the group as an AGENT. The idea here is that the contexts included in these sentences might describe situations in which we (perhaps tacitly) assume only AGENTS are typically involved. The context, combined with the (perhaps tacit) assumption that it describes a situation that (more or less) exclusively involves AGENTS, disposes us to categorize the entity as an AGENT, which then disposes us to attribute to the entity the capacity for phenomenal consciousness.[23] On this account, the reason that attributions of mental states to groups receive lower ratings in the absence of context is because we do not instinctively consider groups to be AGENTS (because they typically do not display certain relevant features). On the interpretation I prefer, the reason that attributions of mental states to groups receive higher ratings in the presence of context is because the context manages to sneak the group through the side door of our AGENT category.[24] Individuals, on the other hand, are unaffected by the presence or absence of context, because we already instinctively categorize them as AGENTS; there simply is no need to sneak them in through the side door.

As I have been interpreting the data, the story behind the "context effect" is that the addition of contextual information sneaks groups in through the "side door" to Agency. There is, though, an alternative interpretation of the data that one might offer, particularly if one is sympathetic to recent work in the fields of collective intentionality or extended cognition.[25] On this interpretation, insofar as my story takes the withholding of consciousness to groups as the "norm" that *adding* context somehow overrides, my

---

[23] On the other hand, we might think that what's happening is that certain contexts call for less stringent versions of the standard phenomenological notions. That is, we might think that 'winning the Super Bowl', for instance, is a context in which we apply the notion of 'feeling proud' more loosely than in other contexts; thus, the folk are willing to attribute 'feeling proud' to the group entity, 'the Giants', because they are using a more inclusive notion of 'feeling' that accommodates groups even though groups typically don't count as AGENTS. The present data is consistent with both interpretations, though I prefer the former explanation for reasons I do not have room to discuss here. That said, I'm open to reconsidering my position should new empirical data support the latter explanation.

[24] On yet another alternative, we might think that there is a separate cognitive process involving a distinct conceptual category, GROUP-AGENT, which is associated with a more austere constellation of behaviors. That is, there may be another set of features that triggers us to categorize a thing as a GROUP-AGENT, which prepotently disposes us to some other constellation of behaviors. This constellation, then, would not include the attribution of phenomenal consciousness, though it likely would also not preclude it. While this alternative is certainly within the logical space of explanations, it strikes me as unnecessarily populating the cognitive system. If the two are equal in explanatory power, I see no need to posit additional concepts/categories; thus, unless there emerges some additional explanatory force out of this alternative explanation, I prefer the more parsimonious position.

[25] Thanks to Mike Bruno for pointing out this alternative interpretation (in e-mail correspondence).

story gets things perfectly backwards. We might tell the story, instead, by beginning with the claim that normal, everyday conversation is rife with attributions (presumably, of consciousness) that include context, which is why subjects rate attributions WITH context so high. Given the overwhelming abundance of contextualized attributions in everyday conversation, we should consider attributions of consciousness (supposedly to both groups and individuals) as the "norm", and it is only when we take away contextual information that sentences deviate from the norm and start to sound so weird. If this story is right, then rather than positing a "context effect", I should be talking instead about the "no context effect" and trying to explain why *removing* context has such an impact on ratings of consciousness attributions to groups and *only* to groups.

This objection, I think, runs up against what I pointed out earlier as the third particularly interesting feature in the data. Namely, there is the fact that even when group attributions of psychological states *include* context, they are still rated significantly less 'natural' than individual attributions of psychological states. If things were such that group attributions of consciousness (so long as they include context) were the "norm", then group attributions WITH context should be rated just as highly as individual attributions of consciousness (WITH or WITHOUT context). They are not. That is, if the standard baseline process were such that we naturally ascribe consciousness to groups (and include context when we do so), then we should observe a similarity, perhaps even a strong correlation, between ratings of individual attributions and ratings of group attributions when both include context. We do not. Given the observed difference between group attributions and individual attributions, it seems to me that the most natural interpretation of the data is that adding context partly overrides (or sneaks past) the default position of withholding consciousness from groups (as described above).

## 6 Conclusion/Future Research

The AGENT-based story told above lends itself to two natural, empirical predictions. First—assuming we can describe a context in this way—presenting subjects with mental state attributions to groups in sentences whose contextual information is maximally AGENT-specific (i.e., roughly, the information

describes a context that is exclusive, or nearly so, to AGENTS[26]) should produce ratings that correlate strongly with ratings of similar attributions to individuals. Second, given that mental state attributions to groups would have to be routed through a "side door" inferential process (the AGENT categorization requires something like inferential reasoning), while mental state attributions to individuals are processed more directly (the AGENT categorization is more automatic), we should expect to see a difference in the time it takes participants to rate the two types of sentences. Thus, the model predicts, a reaction-time study should return a significant lag in affirmative responses for psychological attributions to groups compared to those same attributions to individuals.

We might also expect to see a developmental difference in judgments of attributions of both phenomenal and nonphenomenal states to groups. Given that young children have had fewer opportunities to develop the idea of AGENT-specific contexts (that is, less exposure to situations that are recognized as always being inhabited by AGENTS), it should be harder to sneak groups through this "side door" for AGENCY, thereby making children less likely to be "tricked" into attributing to groups psychological states. As they develop and are exposed to various contexts, entities, social schemas, etc., the more their cognitive processes will be susceptible to context effects like the rest of us. Thus, a study of children's judgments of mental state attributions to groups should reveal, according to the Agency model, a resistance to context effects in very young children and a diminishing resistance as development progresses.

The purpose of this paper has not been to show that the folk *do not* distinguish between phenomenal and nonphenomenal mental states; nor has it been to prove that the folk lack any concept that approximates the philosopher's notion of 'phenomenal consciousness'. My objective is primarily (and simply) to put the empirical onus back on those who would claim that the folk treat phenomenal attributions differently than nonphenomenal attributions, that commonsense psychology distinguishes (even tacitly) between phenomenal and nonphenomenal mental states, or that the folk possess a concept similar to the philosophical notion of 'phenomenal consciousness'. Shifting the burden of proof back upon Biased shoulders might also frustrate those who wish to relate, in some *essential* or *fundamental* way, folk

[26] The trick will be in coming up with contexts that are unique to AGENTS but are not also unique to *individuals*.

attributions of phenomenal consciousness, but *not* attributions of nonphenomenal states, with moral considerations. Given the present data, those theorists who wish to fundamentally or essentially relate moral judgments with attributions of phenomenal consciousness will need either to provide empirical evidence for thinking that the folk *do* distinguish between phenomenal and nonphenomenal states, or to offer some principled reason for doubting that nonphenomenal attributions are also similarly related to moral considerations. Again, I do not presume to have demonstrated the necessary failure of relating moral judgments with attributions of phenomenal (but not nonphenomenal) states. Instead, I take myself merely to have presented an empirical challenge to such theories, a challenge that calls for some explanation (or refutation) from those theories' proponents.

**References:**


Arico, A. 2007. Should Corporate Consciousness be De-Regulated: A Critique of Knobe & Prinz's 'Intuitions About Consciousness. Poster presented at the Annual Meeting of the Society of Philosophy and Psychology, York University, 14-17 June 2007.

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

Augustine. 1995. *Against the Academecians; The Teacher*. King, P. (trans.). Indianapolis: Hackett.

Bloom, P. and Veres, C. 1999. The perceived intentionality of groups. *Cognition*, 71, B1-B9.

Gallagher, S. 2005. *How the Body Shapes the Mind*. Oxford: Clarendon Press.

Goldman, A. 1989. Interpretation Psychologized. *Mind and Language*, 4 pp. 161-185; reprinted in Davies, M. and Stone T. (eds.). 1995. *Folk Psychology: The Theory of Mind Debate.* Oxford: Blackwell.

Gopnik, A. and Meltzoff, A. 1997. *Words, Thoughts and Theories* Cambridge MA: MIT Press.

Gordon, R. 1986. Folk Psychology as Simulation. *Mind and Language*, 1 pp. 158-171; reprinted in Davies, M. and Stone T. (eds.). 1995. *Folk Psychology: The Theory of Mind Debate.* Oxford: Blackwell.

Gray, H., Gray, K., and Wegner, D. 2007. Dimensions of Mind Perception. *Science*, 315 (5812) p. 619.

Heal, J. 1986. Replication and Functionalism. in *Language, Mind, and Logic*. Butterfield, J. (ed.). Cambridge: Cambridge University Press; reprinted in Davies, M. and Stone T. (eds.). 1995. *Folk Psychology: The Theory of Mind Debate.* Oxford: Blackwell.

Heider, F. and Simmel, M. 1944. An experimental study of apparent behavior. *American Journal of Psychology.* 57 pp. 243–259.

Huebner, B. Forthcoming. Commonsense Concepts of Phenomenal Consciousness: Does Anyone Care About Functional Zombies?

Huebner, B., Bruno, M., and Sarkissian, H. Manuscript. What Does the Nation of China Think About Phenomenal States?

Johnson, S. 2003. Detecting Agents. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 358 (1431) pp. 549-559.

Johnson, S., Shimizu, Y., and Ok, S-J. 2007. Actors and actions: The role of agent behavior in infants' attribution of goals. *Cognitive Development*, 22 (3) pp. 310-322.

Knobe, J. 2003. Intentional Action and Side Effects in Ordinary Language. *Analysis*, 63 pp. 190-3.

Knobe, J. and Prinz, P. 2008. Intuitions about consciousness: Experimental studies. *Phenomenology and the Cognitive Sciences*, 7 (1) pp. 67-83.

Mill, J. S. 1865. *An Examination of Sir William Hamilton's Philosophy, and of the Principal Philosophical Questions Discussed in His Writings*. London: Longman, Green, Longman, Roberts & Green.

Reid, T. 1785. *Essays on the Intellectual Powers of Man*. Edinburgh: Printed for John Bell and G.G.J. & J. Robinson.

Robbins, P. and Jack, A. 2006. The Phenomenal Stance. *Philosophical Studies*, 127 pp. 59-85.

Sellars, W. 1956. *Empiricism and the Philosophy of Mind*. In Feigl, H. and Scriven, M. (eds), *Minnesota Studies in the Philosophy of Science* vol. 1. Minneapolis: University of Minesota Press.

Sytsma, J. and Machery, E. Forthcoming. How to study folk intuitions about phenomenal consciousness. *Philosophical Psychology*, forthcoming.

Woodward, A. L. 1998. Infants selectively encode the goal object of an actor's reach. *Cognition*, 69 pp. 1-34.

Zahavi, D. 2005. *Subjectivity and self-hood: Investigating the first-person perspective*. Cambridge, MA: MIT Press.

**Appendix A: Stimulus Items**

Group Phenomenal WITHOUT Context:

McDonalds is feeling upset.

The Art History Department is now vividly imagining.

The Giants felt proud.

The sorority is feeling some regret.

The Housing Authority is feeling angry.

Canada's Travel Bureau is experiencing a sudden urge.

Jamba Juice was experiencing great joy.

The Board of Trustees was feeling relieved.

The F.B.I. is feeling pressure.

The battalion is feeling excruciating pain.

The consulting firm is feeling depressed.

Cisco Systems is feeling exhausted.

Group Phenomenal WITH Context:

McDonalds is feeling upset about the court's recent ruling.

The Art History Department is now vividly imagining a purple square.

The Giants felt proud after they won the Super Bowl.

The sorority is feeling some regret about the recent decision.

The Housing Authority is feeling angry about the foreclosures.

Canada's Travel Bureau is experiencing a sudden urge to pursue internet advertising.

Jamba Juice was experiencing great joy after the promotion.

The Board of Trustees was feeling relieved once the audit was complete.

The F.B.I. is feeling pressure to apprehend terrorist operatives.

The battalion is feeling excruciating pain from the attack.

The consulting firm is feeling depressed about being dumped.

Cisco Systems is feeling exhausted after a rough day on the Stock Exchange.

Group  Nonphenomenal WITHOUT Context:

McDonalds is upset.

The Art History Department is now strongly considering.

The Giants were proud.

The sorority is having some regrets.

The Housing Authority is angry.

Canada's Travel Bureau is having a sudden urge.

Jamba Juice had great joy.

The Board of Trustees was relieved.

The F.B.I. is under pressure.

The battalion is in excruciating pain.

The consulting firm is depressed.

Cisco Systems is exhausted.

Group Nonphenomenal WITH Context:

McDonalds is upset about the court's recent ruling.

The Art History Department is now strongly considering a purple square.

The Giants were proud after they won the Super Bowl.

The sorority is having some regrets about the recent decision.

The Housing Authority is angry about the foreclosures.

Canada's Travel Bureau is having a sudden urge to pursue internet advertising.

Jamba Juice had great joy after the promotion.

The Board of Trustees was relieved once the audit was complete.

The F.B.I. is under pressure to apprehend terrorist operatives.

The battalion is in excruciating pain from the attack.

The consulting firm is depressed about being dumped.

Cisco Systems is exhausted after a rough day on the Stock Exchange.

Individual Phenomenal WITHOUT Context:

Donald is feeling upset.

Sasha is now vividly imagining.

Manning felt proud.

Alison is feeling some regret.

Gillian is feeling angry.

Leslie is experiencing a sudden urge.

Charli was experiencing great joy.

Thomas was feeling relieved.

Mueller is feeling pressure.

Steve is feeling excruciating pain.

Michael is feeling depressed.

Kevin is feeling exhausted.

Individual Phenomenal WITH Context:

Donald is feeling upset about the court's recent ruling.

Sasha is now vividly imagining a purple square.

Manning felt proud after they won the Super Bowl.

Alison is feeling some regret about the recent decision.

Gillian is feeling angry about the foreclosures.

Leslie is experiencing a sudden urge to pursue internet advertising.

Charli was experiencing great joy after the promotion.

Thomas was feeling relieved once the audit was complete.

Mueller is feeling pressure to apprehend terrorist operatives.

Steve is feeling excruciating pain from the attack.

Michael is feeling depressed about being dumped.

Kevin is feeling exhausted after a rough day on the Stock Exchange.

Individual Nonphenomenal WITHOUT Context:

Donald is upset.

Sasha is now strongly considering.

Manning was proud.

Alison is having some regrets.

Gillian is angry.

Leslie is having a sudden urge.

Charli had great joy.

Thomas was relieved.

Mueller is under pressure.

Steve is in excruciating pain.

Michael is depressed.

Kevin is exhausted.

Individual Nonphenomenal WITH Context:

Donald is upset about the court's recent ruling.

Sasha is now strongly considering a purple square.

Manning was proud after they won the Super Bowl.

Alison is having some regrets about the recent decision.

Gillian is angry about the foreclosures.

Leslie is having a sudden urge to pursue internet advertising.

Charli had great joy after the promotion.

Thomas was relieved once the audit was complete.

Mueller is under pressure to apprehend terrorist operatives.

Steve is in excruciating pain from the attack.

Michael is depressed about being dumped.

Kevin is exhausted after a rough day on the Stock Exchange.