# Do you believe that aliens feel pain? An empirical investigation of mental state attributions

Gregory Johnson[1] and Alana Knowles[2]

[1]Department of Philosophy & Religion, Mississippi State University
[1]gregory.johnson@msstate.edu
[2]Mississippi State University

On what basis do we attribute phenomenal states to others? One answer, defended by John Stuart Mill, appeals to an analogy between ourselves and the similar bodies and actions of others (1865, p. 208). Despite its intuitive plausibility, this position is often rejected (Arico et al., 2011; Buckwalter & Phelan, 2014; Knobe & Prinz, 2008). In line with Mill's account, we propose that the primary factors used when making phenomenal state ascriptions are the appropriate display of functional and behavioral cues and having bodies with the right kind of physical composition. To test this account, we gave five groups of participants a vignette followed by three to six questions. For four of the groups, the vignette described an alien-human encounter and the participants had to judge the likelihood (on a 7-point scale) that the alien had a non-phenomenal state (a belief) and the likelihood that it could have a phenomenal state (pain). The fifth group, as a control, read a vignette about a similar interaction between two humans. We found that, as appropriate functional and behavioral cues and then humanoid features are added to the alien, people are more willing to attribute a phenomenal state to it. Attributions of the non-phenomenal state are primarily dependent on the appropriate functional and behavioral cues, not on having humanoid features.

*keywords: phenomenal consciousness, non-phenomenal states, other minds, experimental philosophy, alien minds*

This is the penultimate version of a paper that will appear in *Cognition, Brain, Behavior*.

## 1 Introduction

Gray et al. (2007), and Knobe and Prinz (2008) launched a cottage industry devoted to experimental investigations of people's attributions of mental states to non-humans—corporations, robots, ghosts, chimpanzees, God, and others. Both found that people make a distinction between intentional and phenomenal mental states and, in some cases, will attribute one but not the other to an entity. In Knobe and Prinz's (2008) study, the participants were willing to attribute intentional states (e.g., deciding, intending, believing) to groups, but they were unwilling to do the same for phenomenal states (e.g., "experiencing great joy," "feeling excruciating pain"). The participants in Gray et al.'s (2007) study, meanwhile, attributed agency (meaning roughly, *having intentional states*) but not experience (meaning roughly, *having phenomenal states*) to God and a robot. At the same time, they attributed experience but little or no agency to a frog, a human fetus, a man in a persistent vegetative state, a dog, a chimpanzee, and a baby. And, not surprisingly, the participants attributed roughly equal degrees of agency and experience to themselves or another human—as well as to a dead woman, although the actual level of agency and experience attributed to the dead woman was much less than it was for themselves or another living human.[1]

Subsequent work has investigated whether people are actually attributing intentional states to groups and not just to the members of each group (Phelan et al., 2013), whether people attribute mental states to the dead and to persons in persistent vegetative states (Gomes & Parrott, 2015; Gray et al., 2011), if non-experts have the same unified concept of phenomenal consciousness as do philosophers (Buckwalter & Phelan, 2014; Fiala et al., 2014; Sytsma & Machery, 2010; Sytsma & Ozdemir, 2019), how non-experts divide up the mental space if not in terms of *phenomenal* and *non-phenomenal* (Malle, 2019; Weisman et al., 2017), and the relationship between attributions of agency or free will and attributions of consciousness (Arico et al., 2011; Björnsson & Shepherd, 2020; Nahmias et al., 2020; Shepherd, 2015). The question that grew out of this work and interests us, however, is what is the basis on which people attribute phenomenal mental states to other entities?[2] That is, when people do attribute such a mental state to an entity, what are the minimal grounds that they use to do so?

We propose that the primary factors used when making phenomenal state ascriptions are the display of relevant functional and behavioral cues and having the right kind of physical composition. To test this model, we gave five groups of participants a vignette followed by three to six questions. For four of the groups, the vignette described an alien-human encounter and the participants had to judge the likelihood (on a 7-point scale) that the alien had a non-phenomenal mental state (a belief) and the likelihood that it could have a phenomenal state (pain). The fifth group, as a control, read a vignette about a similar interaction between two humans.

## 1.1 The embodiment and functionalist hypotheses

Buckwalter and Phelan (2014) consider two explanations for when and how people attribute phenomenal states to other entities. According to the *embodiment hypothesis*,

[1] Gray et al.'s *agency* and *experience* are their own labels for the two groupings that emerged from their survey data. For the most part, these categories line up with the standard philosophical distinction between *phenomenal* and *intentional*. *Experience* consists of mental states that they define using the terms *feeling* or *experiencing*: embarrassment, fear, hunger, joy, pain, pleasure, pride, and rage—plus consciousness (i.e., being "capable of having experiences and being aware of things"), desire (the ability to hope or long for things), and the ability to have personality traits. *Agency* consists of the ability to communicate, to recognize emotions in others, to remember, to tell right from wrong, to plan, to exercise self-restraint, and to think.

[2] Of course, as some of the work just mentioned suggests, non-philosophers may not have as broad a notion of phenomenal consciousness as is found in the philosophical literature. This need not concern us here. Our study investigates "feeling pain" and attributions of such experiences still need an explanation.

> Unified *biological embodiment* is a major psychological factor that cues ordinary attribution of experiences, feelings, emotions, and so on, to other entities. The strongest version of this view is that phenomenal attribution requires biological embodiment. Weaker versions focus on relative levels of attribution, claiming that phenomenal attributions are more likely to be cued as an entity's biological body becomes more salient. (2014, p. 46)

In contrast, the *functionalist hypothesis* has it that "functional information – information about goals, desires, and so forth, of an entity – tends to cue phenomenal state ascription independently of whether the entity has a unified biological body" (2014, p. 50).

Buckwalter and Phelan (2014) provide some support for the functionalist hypothesis with a series of experiments in which participants were given a story about either a human, a ghost, or an "eternally disembodied spirit." They found that participants were equally willing to attribute feeling happy, feeling angry, and feeling sad to all three types of entities. For instance, in one experiment, two groups of participants read a story about a man named Bob who, at the beginning of the story, gets divorced from Melissa. Bob then takes some steps to turn their son against his mother, who, in the meantime, has begun a new relationship. At a pivotal moment in the story that was given to one of the groups, Bob is killed in a car accident. He, however, "emerges from his dead body as a ghost" (2014, p. 53). In the story given to the other group, Bob is not in a car accident and isn't killed. Then, either as a ghost or as a human, Bob places some photographs of Melissa on a date where their son will find them. Following the story, participants rated their level of agreement with this statement on a 7-point scale: "As Bob moves the pictures into place, he feels angry at Melissa for beginning a new relationship." Mean scores for both groups were virtually identical (human condition: $M$ = 6.06, $SD$ = 0.91, ghost condition: $M$ = 6.11, $SD$ = 1.19).

Knobe and Prinz (2008), meanwhile, defend a version of the embodiment hypothesis. They propose that "information about physical constitution plays a special role in those ascriptions that require phenomenal consciousness – a role that it does not play in other kinds of mental state ascription" (2008, p. 70). They focus on group agents, in particular, corporations, and find that, while people are willing to ascribe non-phenomenal states to Microsoft or "Acme Corp," they are unwilling to do the same with phenomenal states. In the main experiment supporting their hypothesis, they gave participants a list of sentences that attributed non-phenomenal and phenomenal states to a corporation – they used, for example, "Acme Corp. believes that its profit margin will soon increase," "Acme Corp. intends to release a new product this January," "Acme Corp. is now experiencing great joy," and "Acme corp. is feeling excruciating pain" (pp. 74–75). Participants rated each sentence from 1 ("sounds weird") to 7 ("sounds natural"). They found that the mean scores for sentences describing Acme Corp as *knowing*, *believing*, *intending*, *wanting*, and *deciding* were between 5.2 and 6.6, while the mean scores for sentences describing Acme Corp as *feeling excruciating pain*, *getting depressed*, *vividly imagining*, and *experiencing* *great joy* were between 2.1 and 4.7.[3]

Knobe and Prinz also investigated the basis on which people attribute phenomenal states (2008, pp. 75–77). One possibility that they consider is that judgements about whether an entity has phenomenal states is based on that entity's similarity to humans. They write,

> Subjects start out with the premise that human beings have phenomenal consciousness. Then, when they are wondering whether some other sort of agent has phenomenal consciousness, they simply ask whether its physical constitution is sufficiently similar to that of human beings. Since the physical constitution of a corporation is extremely unlike that of a human being in numerous respects, subjects conclude that corporations do not have phenomenal consciousness. (2008, p. 76)

On the other hand, they also consider this possibility:

> Perhaps subjects are not thinking at all about similarity to human beings. Perhaps they are applying a far more specific restriction on constitution (say, a restriction against agents that are composed of other agents). On this latter view, people might be willing to ascribe phenomenal states to agents that are very, very different from us—just as long as those agents do not violate the specific restriction. (2008, p. 76)

To test these competing hypotheses, they gave participants this vignette:

> Once there was a powerful sorceress. She came upon an ordinary chair and cast a spell on it that endowed it with a mind. The chair was still just made of wood, but because of the magic spell, it could now think complex thoughts and form elaborate plans. It would make detailed requests to the people around it, and if they didn't do everything just as it wanted, it would start complaining. People used to call it the Enchanted Chair. (2008, p. 76)

The participants then answered this question: "Can the Enchanted Chair *feel happy* or *sad*?" The same participants were also given a brief description of a corporation—although not one that had been enchanted by a sorceress—and asked, "Can Acme Corp. *feel happy* or *sad*?" Participants answered by selecting a number from 1 (disagree) to 7 (agree). The

[3] These results were partly replicated by Huebner et al. (2010) who compared responses to similar questions from students at a university in the United States and students at the Chinese University of Hong Kong—although, for the questions about the mental states of groups, they used sentences that could more easily be associated with actual groups and states of affairs (e.g., "The Ming Dynasty felt relief after the rebellion was quelled," "Denmark feels embarrassed about losing the war"). Huebner et al. found that both the U.S. and the Chinese students thought that ascribing a phenomenal state to an individual sounded more natural than ascribing a phenomenal state to a group. There was a much smaller difference, however, between how natural the Chinese students judged ascriptions of phenomenal states to individuals ($M$ = 4.69) and ascriptions of phenomenal states to groups ($M$ = 4.06) than there was for the U.S. students ($M$ = 5.74 and $M$ = 3.44 respectively).

average score for the question about the chair was 5.6, while, for the question about the corporation, it was 1.8.

This doesn't tell us what the basis for attributing phenomenal states might be, but it suggests that the range of beings to which such states are attributed includes enchanted chairs but not corporations. So, while Knobe and Prinz (2008) maintain that composition is a relevant feature when deciding whether to attribute a phenomenal state to an entity, they don't try to specify what kind of composition is required. Based on their results, it can't be an agent, like a corporation, that is composed of other agents, but neither does the entity have to be especially similar to a human.

An alternative version of the embodiment hypothesis – the one that Knobe and Prinz (2008) first considered – rests on an analogical argument. You know that you have phenomenal states. When you observe that others are, in many outward respects, similar to you, you conclude that they have similar inner experiences. The classic statement of this view is given by John Stuart Mill. He writes,

> By what considerations am I led to believe, that there exist other sentient creatures; that the walking and speaking figures which I see and hear, have sensations and thoughts, or in other words, possess Minds? …I conclude that other human beings have feelings like me, because, first, they have bodies like me, which I know, in my own case, to be the antecedent condition of feelings; and because, secondly, they exhibit the acts, and other outward signs, which in my own case I know by experience to be caused by feelings. (1865, p. 208)

Regarding humans, Mill may be right, although it's possible that, for the majority of cases, we just follow the rule that other humans have phenomenal states without making any sort of inference. But setting this worry aside, we find Mill's proposal attractive. (For more recent defenses of this position, see Hill [1991] and Hyslop [1995].)

As a starting point, we propose that the primary features used for inferring the presence or absence of phenomenal states are functional and behavioral cues and physical composition. The functional and behavioral cues are observed when a creature interacts with its environment in an appropriate way or shows certain abilities that humans typically possess (e.g., abilities to plan and organize). And the entity's physical composition should be organic with a shape that is relatively similar to that of a human's (e.g., possessing arms, legs, mouth, eyes, as well as a digestive system, brain, etc.). For this study, we describe a living creature and vary, across conditions, its body's shape and form, its sensory organs, and whether it displays intentional behaviors (greeting someone and then helping her).

# 2 Method

## 2.1 Participants

A total of 359 adults, all of whom were in the United States, participated in the study, which was administered using Mechanical Turk (www.mturk.com). Participants were paid $0.60 for participating in the first four groups and $0.75 for participating in the human control condition. (The differences between the first four conditions and the human control condition are explained below.) Participants could only be in one group. After removing responses from participants who didn't pass the comprehension or time conditions (described below), 296 participants were included in the analysis (43 percent female, average age: 35, age range: 18 to 70). The study was reviewed and approved by the Mississippi State University institutional review board (Protocol ID: IRB-18-483).

## 2.2 Materials

### 2.2.1 Vignettes and questionnaires

Five groups of participants were given different vignettes to read; each vignette was followed by either three or six questions. Four of the vignettes described a human-alien encounter.

1. **alien control condition** Dr. Jane Stevenson is a scientist who works in the physics department at a large university. One day, when she returns from lunch, she finds what appears to be an alien in her office. The creature looks like a floating blob of mercury. It is not solid, and it has no discernible face. Jane waves her hand and says, "Hello?" The creature only continues to hover in the same spot.

2. **math control condition** Dr. Jane Stevenson is a scientist who works in the physics department at a large university. One day, she spends most of the morning working on a difficult math problem on the white board in her office. Later, when she returns from lunch, she finds what appears to be an alien in her office. The creature looks like a floating blob of mercury. It is not solid, and it has no discernible face. Jane waves her hand and says, "Hello?" The creature only continues to hover in the same spot. Jane looks around her office. Suddenly she notices a new equation written on the board where she had been working:

   $$\mathcal{L}_H = [(\delta_\mu - igW_\mu^\alpha \tau^\alpha - i\tfrac{1}{2}g'B_\mu)\theta]^2$$

   The alien turns from her and adds $+\, \mu^2\, \phi^t\, \phi - \lambda(\phi^t\, \phi)^2$ to the end of the equation by slowly moving a part of its body over the board.

3. **functional and behavioral cues condition** Dr. Jane Stevenson is a scientist who works in the physics department at a large university. One day, when she returns from lunch, she finds what appears to be an alien in her office. The creature looks

like a floating blob of mercury. It is not solid, and it has no discernible face. Jane waves her hand and says, "Hello?" The creature responds by moving up and down and making a brief sound. Jane is overwhelmed and feels like she is about to faint. As she starts to fall, the alien rushes to her, wraps part of its body around her and guides her into a chair.

4. **all cues condition** Dr. Jane Stevenson is a scientist who works in the physics department at a large university. One day, when she returns from lunch, she finds what appears to be an alien in her office. The creature is a little bit taller than she is, it's standing upright on two legs, and it has two arms, a torso, neck, and head. But its eyes are large and clear, and Jane can see something pulsating behind them. It has a small mouth. It's hairless, and its skin is a pale orange and green. It has three long and slender fingers on each hand.
   Jane waves her hand and says, "Hello?" The creature responds by raising its hand and emitting a brief tone. Jane is overwhelmed and feels like she is about to faint. As she starts to fall, the alien rushes to her, grabs her and guides her into a chair.

Each one of these four vignettes was followed by these three questions:

(*a*) Is there a creature, which is probably an alien, in Jane's office?
(*b*) Does the alien believe that it is in a room with Jane?
(*c*) If Jane cut the alien with a sharp knife, would the alien feel pain?

Participants were instructed to answer the questions using a 7-point scale where 1 means "clearly no," 4 means "not sure," and 7 means "clearly yes." The final vignette described a human-human encounter.

5. **human control condition** Dr. Jane Stevenson is a scientist who works in the physics department at a large university. One day, when she returns from lunch, she finds a man waiting in her office. Jane isn't sure who this person is, but she says, "Hello."
   The man responds, "Oh, sorry, I thought this was Dr. Moore's office."
   Jane is about to explain that Dr. Moore's office is on a different floor, but suddenly she feels as if she is about to faint. As she starts to fall, the man rushes to her, grabs her and guides her into a chair.

This vignette was followed by six questions. Participants answered four of them (a, b, d, and f) using the 7-point scale. They gave written answers for (c) and (e).

(*a*) Is there a man in Jane's office?
(*b*) Does this man believe that he is in a room with Dr. Moore?
(*c*) Please explain your answer for (b).
(*d*) Does this man believe that he is in a room with a woman who is now in a chair?
(*e*) Please explain your answer for (d).
(*f*) If Jane cut this man with a sharp knife, would the man feel pain?

### 2.2.2 Explanations of the conditions

In the first condition (the alien control condition), the alien's behavior, apparent composition, and body-type were such that our model predicted that participants would not attribute mental states—or at least not phenomenal states—to it. The alien in the third vignette (the functional and behavioral cues condition) had the same body-type as in the first, but it displayed some relevant functional abilities. And the alien in the fourth vignette (the all cues condition) displayed the same functional abilities as in the previous condition but had a more humanoid appearance. Our expectation was that participants in each of these respective groups (the first, third, and fourth) would show more confidence about attributing mental states to the alien than the participants in the previous group. (That is, the participants in the third group would be more confident than those in the first, and those in the fourth would be more confident than those in the third.)

The purpose of the math control condition was to distinguish between functional abilities that are the basis for attributing mental states to a creature and those that are not. Our expectation was that only displaying an ability for complex math would not be the basis for attributing mental states to the alien, and so the scores in this condition would be closer to those in the alien control conditional than to those in any of the other conditions. Finally, as an additional control condition, a fifth group (the human control condition) read a vignette in which a human replaced the alien. Here, we expected the scores to reflect the maximum confidence that one can have about attributing mental states—whether phenomenal or non-phenomenal—to another being.

Creating the survey for the *human control* condition presented some unforeseen difficulties. Asking "Does this man believe that he is in a room with Jane?" is ambiguous between *Does he believe that he is in a room with a woman (who happens to be Jane)?* and *Does he believe that he is in a room with Dr. Jane Stevenson?* And asking *Does this man believe that he isn't in a room with Dr. Moore?* seems to be a belief that some people wouldn't ascribe to the man in the vignette (i.e., a belief about a negative state of affairs). So, questions (b) and (c) were included to help participants begin thinking about the question in (d), and we did not use the results from (b) and (c).

### 2.2.3 Exclusions

For the four groups that had an alien in the vignette, we rejected the results from participants who answered, "Is there a creature, which is probably an alien, in Jane's office?" with a 4 or lower. Results were also rejected if a participant finished the survey in less than one minute. All participants in the human control condition answered the comprehension question "Is there a man in Jane's office?" with a 7, but as an additional evaluation of comprehension, we both looked at the written answers for (e) before looking at any of the other results. We gave each participant's response a score from 0 (the person definitely did not understand the question) to 5 (the person definitely did understand

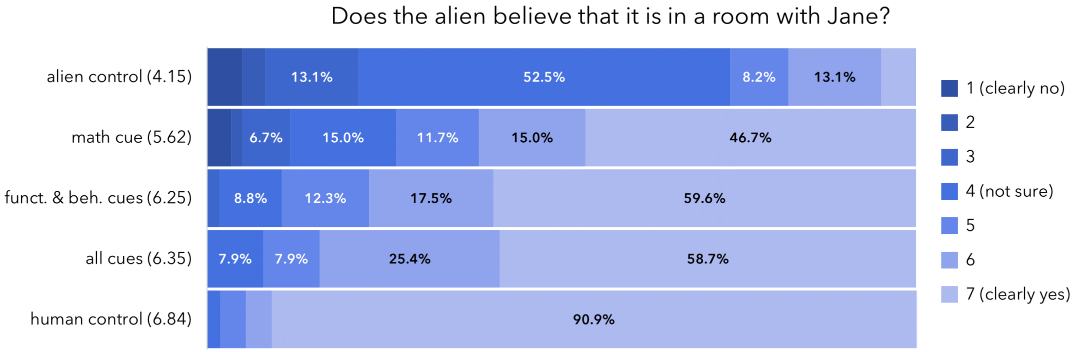

**Figure 1:** The mean scores (in parentheses) and the distribution of scores for the belief question. The first four groups answered the question "Does the alien believe that it is in a room with Jane?" The human control group answered "Does this man believe that he is in a room with a woman who is now in a chair?" The median scores for each group (in the order in which they are listed) were 4, 6, 7, 7, and 7. The following data labels are not shown in the chart. Alien control: (1) 4.9%, (2) 3.3%; (7) 4.9%; math cue: (1) 3.3%, (2) 1.7%; functional and behavior cues: (3) 1.8%; human control: (4) 1.8%, (5) 3.6%, (6) 3.6%. Values not given are 0.0%.

the question). If one or both of us gave the answer a 0 or a 1, then we did not use that participant's answers for any of the questions.

# 3 Results

## 3.1 Belief question

For the four groups that read a vignette about an alien, the question about having a belief was *Does the alien believe that it is in a room with Jane?* For the human control condition, it was *Does this man believe that he is in a room with a woman who is now in a chair?* Participants answered by selecting a number from 1 (meaning "clearly no") to 7 (meaning "clearly yes"). The mean score for each group and the distribution of scores are given in figure 1. A one-way ANOVA was performed to compare the effect of the group (alien control, math control, etc.) on the mean score for the attribution of belief. There was a statistically significant difference in the mean scores at the $p < .05$ level for the five groups, $F(4, 291) = 45.95$, $p < .001$.

Post hoc comparisons were made using the Tukey HSD test. Participants in the alien control condition were the least willing to attribute the belief to the alien, and the mean for the alien control condition ($n = 61$, $M = 4.15$, $SD = 1.33$) was significantly different than the means for all of the other groups. The mean for the math cue condition ($n = 60$, $M = 5.62$, $SD = 1.67$) was also significantly different than the other conditions. The differences between the means for the functional and behavioral cues condition ($n = 57$, $M = 6.25$, $SD = 1.09$), the all cues condition ($n = 63$, $M = 6.35$, $SD = .94$), and

| | math cue | functional & behavioral cues | all cues | human control |
|---|---|---|---|---|
| alien control | .000 (0.98) | .000 (1.72) | .000 (1.92) | .000 (2.59) |
| math cue | | .035 (0.44) | .006 (0.55) | .000 (0.96) |
| functional & behavioral cues | | | .989 (0.10) | .066 (0.68) |
| all cues | | | | .172 (0.62) |
| $p \leq 0.05$ | | | | |

**Table 1:** Significance levels for the difference in means for each pair-wise comparison and effect sizes (in parentheses) for the belief question.

human control condition ($n$ = 55, $M$ = 6.84, $SD$ = .57) were not significant. The pairwise significance levels for the differences between means are given in table 1.[4]

### 3.2 Pain question

For the four alien conditions, the question about pain was *If Jane cut the alien with a sharp knife, would the alien feel pain?* In the human control condition, the question about pain was *If Jane cut this man with a sharp knife, would the man feel pain?* Participants answered using the same 7-point scale. The distribution of scores and the means are given in figure 2. A one-way ANOVA was performed to compare the effect of the group (alien control, math control, etc.) on the mean score for the attribution of pain. There was a statistically significant difference in the mean scores at the $p$ < .05 level for the five groups, $F$(4,291) = 38.6, $p$ < 0.001.

Post hoc comparisons were made using the Tukey HSD test. Again, the mean score for the alien control group was the lowest ($M$ = 4.03, $SD$ = 1.69), and it increased for each group in the expected order: the math cue condition ($M$ = 4.47, $SD$ = 1.65), the functional and behavioral cues condition ($M$ = 5.0, $SD$ = 1.64), the all cues condition ($M$ = 5.86, $SD$ = 1.33), and the human control condition ($M$ = 6.96, $SD$ = .19). The differences between the means were not significant for (1) the alien control condition and math cue condition and (2) the math cue condition and the functional and behavioral cues condition. The other pairwise comparison of means were significantly different (see table 2).

[4] The ANOVA and Tukey HSD test were performed in R. The effect sizes were measured with Cohen's $d$ using $\sqrt{\frac{(n_1-1)sd_1^2+(n_2-1)sd_2^2}{n_1+n_2-2}}$ to compute the pooled standard deviation.

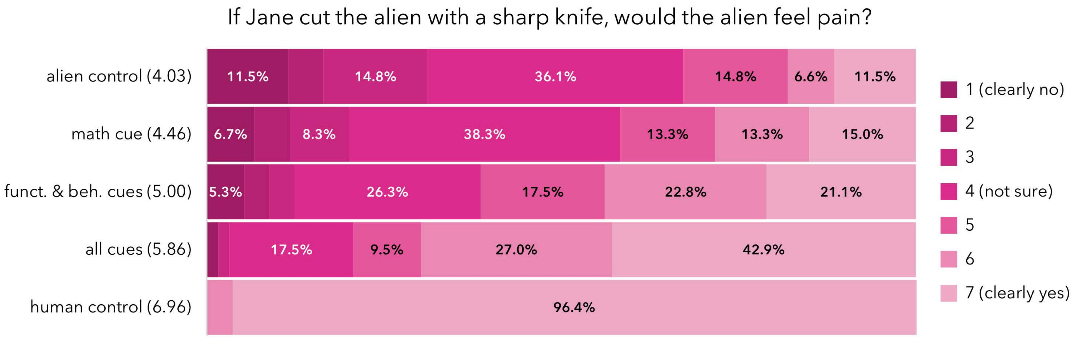

**Figure 2:** The mean scores (in parentheses) and the distribution of scores for the pain question. The first four groups answered the question "If Jane cut the alien with a sharp knife, would the alien feel pain?" The human control group answered "If Jane cut this man with a sharp knife, would the man feel pain?" The median scores for each group (in the order in which they are listed) were 4, 4, 5, 6, and 7. The following data labels are not shown in the chart. Alien control: (2) 4.9%; math cue: (2) 5.0%; functional and behavior cues: (2) 3.5%, (3) 3.5%; all cues: (1) 1.6%, (3) 1.6%; human control: (6) 3.6%. Values not given are 0.0%.

## 4 Discussion

Sometimes, although not always, we attribute phenomenal states to others. We generally do so for other humans, and in some cases, we attribute such states to non-human creatures. We developed and then tested a model that explains when such attributions are made. Consistent with our model, as (1) appropriate functional and behavioral cues and then (2) humanoid features are added to a creature, people are more willing to attribute a phenomenal state (pain) to it. Attributions of a non-phenomenal state (a belief), meanwhile, were only dependent on the appropriate functional and behavioral cues, not on having a certain kind of composition.

| | math cue | functional & behavioral cues | all cues | human control |
|---|---|---|---|---|
| alien control | .455 (0.26) | .002 (0.58) | .000 (1.20) | .000 (2.37) |
| math cue | | .260 (0.32) | .000 (0.93) | .000 (2.08) |
| functional & behavioral cues | | | .010 (0.58) | .000 (1.67) |
| all cues | | | | .000 (1.13) |
| $p \leq 0.05$ | | | | |

**Table 2:** Significance levels for the difference in means for each pair-wise comparison and effect sizes (in parentheses) for the pain question.

## 4.1 Belief question

In the alien control condition, the alien did nothing, and so there were no functional or behavioral cues. As expected, the mean score for this group was almost right at 4, which represented “not sure”. For the last three conditions (the functional and behavioral cues, all cues, and human control), the functional and behavioral cues remained, basically, the same: the alien or human responds to Jane and helps her into a chair. The composition of the agent across these three conditions did, however, vary: from what appears to be a floating blob of mercury, to a humanoid alien, to a human. The alien in the math cues condition, meanwhile, was compositionally the same as the alien in the first and third conditions, but it doesn’t respond when Jane says hello and, for all we know, its abilities are only mathematical.

The most interesting result here is that the functional and behavioral cues that are present in the last three conditions appear to be all that is needed for ascribing an intentional state like *the belief that I am in a room with Jane* to a creature. The changes in composition across those three conditions add little or nothing to people’s willingness to attribute the belief. (Although the means for those three conditions vary, none of the differences between those means are statistically significant.) Thus, assuming that we have the maximum amount of confidence when attributing this belief to the human, the functional and behavioral cues described in the vignettes get us that maximum amount of confidence, even when the agent is a floating blob of mercury. The mean score for the math cue condition falls between the mean for the alien control condition and the functional and behavioral cues condition and is significantly different than both. This suggests that some kind of goal directed behavior is sufficient to make us a little more confident about attributing a belief than we are in the absence of any functional or behavioral cues, but the wrong kind of goal directed behavior keeps that confidence at a modest level.

## 4.2 Pain question

The question about feeling pain shows that attributing a phenomenal state is sensitive to functional and behavioral cues and to composition. Again, we found that the mean for the alien control condition was almost right at 4. Then, for the functional and behavioral cues condition, the all cues condition, and the human control condition, each mean revealed progressively more confidence in the creature being able to experience pain. These results support our version of the embodiment hypothesis. The functional cues on display matter, as demonstrated by the different responses in the alien control condition and the behavioral and functional cues condition. But the attribution of phenomenal states is also sensitive to embodiment. When functional and behavioral cues are held constant (as they are for the functional and behavioral cues, all cues, and the human control conditions), the right kinds of changes in composition made participants more willing to agree that the agent could experience pain.

### 4.3 General discussion

According to our model, people attribute phenomenal states to other agents who display the right kinds of functional and behavioral cues and the appropriate kind of physical composition. This model is supported by the results of our study on the attribution of pain to the aliens described in the vignettes. We want to stress, however, that this is an initial model. Pain is just one example of a phenomenal state. It's a prominent mental state, but it remains to be shown that other phenomenal states will yield similar results.

Also worth noting is that we found, depending on the creature being evaluated, that people occupy one of these three categories:

(*a*) They are not confident about attributing phenomenal or non-phenomenal states to the creature.

(*b*) They are confident about attributing a non-phenomenal state to the creature but are less confident about attributing a phenomenal state to it.

(*c*) They are confident about attributing both non-phenomenal and phenomenal states to the creature.

Hence, in this space of mental state attributions, attributions of phenomenal states are a more selective attribution. But there are probably creatures for which neither (a), (b), nor (c) would be the case. For instance, most people, we imagine, would be willing to attribute feeling pain but not intentional states to some non-primate mammals, a possibility that is suggested by Gray et al.'s results (2007; see also Weisman et al., 2017). Such attributions may not be covered especially well by our model. More work remains to be done, but the current version of our model might be better at explaining phenomenal state attributions when (a), (b), and (c) are the possibilities in play.

Furthermore, Buckwalter and Phelan's (2014) results seem to show that people will attribute phenomenal states (feeling happy, feeling angry, and feeling sad) to ghosts and "eternally disembodied spirits," and Knobe and Prinz's results suggest that people will attribute phenomenal states to an "enchanted chair." These results conflict with our composition requirement. We offer two responses.

First, Gomes and Parrott (2015) point out that, in these kinds of studies, participants may be making attributions based on what is "true only *according to the story*" (p. 1003). These "truths in fiction" could lead people to attribute emotional states to an enchanted chair or to a ghost or an eternally disembodied spirit if they find that the events in the story seem to require making such an attribution. The story that Knobe and Prinz (2008) used all but says that the enchanted chair can get angry when its demands aren't met, and so it's possible—or perhaps likely—that, while participants were agreeing that the chair could feel happy or sad *in the story*, they were not committing themselves to any particular stance on attributing phenomenal states in reality. Similarly, in the version of

Buckwalter and Phelan's (2014) story where Bob dies in a car accident, the ghost version of Bob retains the same goal that Bob had before the accident. Insofar as this goal is motivated by being upset and angry, it is, again, possible that participants were agreeing that the ghost in the story must feel angry without committing themselves to anything further.

Our study may also suffer from some degree of participants answering on the basis of 'true according to the story,' although the scores for the alien control condition and the human control condition show that participants were unsure about attributing mental states to the alien that displayed none of the relevant cues and were very confident about doing so for the human—results that appear to reflect situating the events in reality. Moreover, we guarded against 'true according to the story' affecting our results by putting different versions of an alien in the same story and tracking the different responses that the participants had to each version. The finding that, as the relevant cues are added to the story, people are more willing to make these mental state attributions demonstrates that the cues are driving the attributions.

Second, the inference that is made according to our model may be overridden by other rules or guidelines that people follow for specific types of creatures. As Fiala et al. (2014) explain,

> 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 [i.e., conscious, deliberate] reasoning about robots. If so, then subjects will show significant resistance to attributions of mental states to robots generally. (2014, p. 37)

Similarly, it may be a platitude in our culture that ghosts—like the ones in Buckwalter and Phelan's (2014) study—can have some or all of the same phenomenal states as humans. Moreover, the ghost in Buckwalter and Phelan's story can causally interact with its surroundings, and so participants in their study may have interpreted the creature as having some sort of physical embodiment. In contrast, our model is designed to explain the attributions that people make in the absence of any special rules or guidelines about the phenomenal states of specific types of creatures. People may very well have a rule of thumb that guides them when thinking about aliens, but, even if that is the case, our study demonstrates that, while keeping the type of creature constant across all of the alien conditions, there are certain features that are the basis for making mental state attributions.

## References

Arico, A., Fiala, B., Goldberg, R. F., & Nichols, S. (2011). The Folk Psychology of Consciousness. *Mind & Language*, 26(3), 327–352.

Björnsson, G., & Shepherd, J. (2020). Determinism and attributions of consciousness.

*Philosophical Psychology*, 33(4), 549–568.

Buckwalter, W., & Phelan, M. (2014). Phenomenal Consciousness Disembodied. In J. Sytsma (Ed.), *Advances in Experimental Philosophy of Mind* (pp. 45–74). Bloomsbury Academic.

Fiala, B., Arico, A., & Nichols, S. (2014). You, Robot. In *Current Controversies in Experimental Philosophy* (pp. 31–47). Routledge.

Gomes, A., & Parrott, M. (2015). Epicurean aspects of mental state attributions. *Philosophical Psychology*, 28(7), 1001–1011.

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science*, 315(5812), 619.

Gray, K., Knickman, T. A., & Wegner, D. M. (2011). More dead than dead: Perceptions of persons in the persistent vegetative state. *Cognition*, 121(2), 275–280.

Hill, C. S. (1991). Sensations: A defense of type materialism. Cambridge University Press.

Huebner, B., Bruno, M., & Sarkissian, H. (2010). What Does the Nation of China Think About Phenomenal States? *Review of Philosophy and Psychology*, 1(2), 225–243.

Hyslop, A. (1995). Other minds. Springer.

Knobe, J., & Prinz, J. (2008). Intuitions About Consciousness: Experimental Studies. *Phenomenology and the Cognitive Sciences*, 7(1), 67–83.

Malle, B. F. (2019). How many dimensions of mind perception really are there? In A. K. Goel, C. M. Seifert, & C. Freksa (Eds.), *Proceedings of the 41st Annual Meeting of the Cognitive Science Society* (pp. 2268 - 2274).

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

Nahmias, E., Allen, C. H., & Loveall, B. (2020). When Do Robots have Free Will? Exploring the Relationships between (Attributions of) Consciousness and Free Will. In B. Feltz, M. Missal, & A. Sims (Eds.), *Free Will, Causality, and Neuroscience* (Vol. 338, pp. 57–80). Brill.

Phelan, M., Arico, A., & Nichols, S. B. (2013). Thinking things and feeling things: On an alleged discontinuity in folk metaphysics of mind. *Phenomenology and the Cognitive Sciences*, 12(4), 703–725.

Shepherd, J. (2015). Consciousness, free will, and moral responsibility: Taking the folk seriously. *Philosophical Psychology*, 28(7), 929–946.

Sytsma, J. & Machery, E. (2010). Two conceptions of subjective experience. *Philosophical Studies*, 151, 299 – 327.

Sytsma, J. & Ozdemir, E. (2019). No Problem: Evidence that the Concept of Phenomenal Consciousness is Not Widespread. *Journal of Consciousness Studies*, 26(9–10), 241–256.

Weisman, K., Dweck, C. S., & Markman, E. M. (2017). Rethinking people's conceptions of mental life. *Proceedings of the National Academy of Sciences*, 114(43), 11374–11379.