

On the Relevance of Folk Intuitions: A Reply to Talbot Justin Sytsma and Edouard Machery¹

Abstract: In previous work, we presented evidence suggesting that ordinary people do not conceive of subjective experiences as having phenomenal qualities. We then argued that these findings undermine a common justification given for the reality of the hard problem of consciousness. In a thought-provoking article, Talbot has challenged our argument. In this article, we respond to his criticism.

In previous work (Sytsma and Machery, 2010), we presented evidence suggesting that ordinary people do not conceive of subjective experiences as having phenomenal qualities, and we argued that these findings cast doubt on the reality of the hard problem of consciousness. In a thought-provoking article, Talbot challenges our argument: He holds that the kinds of judgments we reported in our previous work do not cast doubt on the reality of the hard problem of consciousness because they result from an intuitive, unreflective, similarity-based cognitive system (also called “System One”). In this article, we respond to his criticism.

In Section 1, we review briefly the argument put forward in our previous work, before describing the key points of Talbot’s criticism in Section 2. In Section 3, we develop our first line of response to this criticism: Granting for the sake of argument that judgments ascribing mental states (such as the ascription of seeing a color or of a feeling of pain) can result from one of two cognitive systems—System One or System Two—we show that, when System-Two judgments are elicited, people still treat visual states differently from feelings of pain, as we found previously. Section 4 presents our second line of response: Contra Talbot, we show that the judgments reported in our previous work do not result from an intuitive, unreflective, similarity-based cognitive system.

¹ The first author did most of the work for this article. The authors would like to thank Adam Arico, Joshua Knobe, and Mark Phelan for their helpful comments on a previous draft of this article.

1. Two Conceptions of Subjective Experience

In Sytsma and Machery (2010) we presented empirical evidence that ordinary people (viz. people without training in philosophy or in consciousness studies) do not tend to classify mental states as philosophers do. Specifically, while the philosophers we surveyed tended to treat such diverse mental states as seeing red and feeling pain *similarly*, denying that a simple non-humanoid robot can have either, the ordinary people we surveyed tended to treat those states *differently*: Like the philosophers, they denied that the robot feels pain, but in contrast to the philosophers they asserted that the robot sees red.

Having found evidence that ordinary people do not classify mental states as philosophers do, we went on to explore the classification scheme that they do employ. While philosophers tend to classify mental states in terms of whether or not they are phenomenally conscious, we presented evidence suggesting that ordinary people instead tend to classify mental states in terms of whether or not they are thought to have a valence.

On the basis of these findings, we concluded that, in general, ordinary people do not call on the philosophical distinction between phenomenal and non-phenomenal in classifying mental states. And on the basis of this empirical conclusion, we then argued for a philosophical conclusion: We argued that our findings cast doubt on a common justification given for the reality of the hard problem of consciousness. While some philosophers have justified the claim that there is a real problem here by arguing that the existence of phenomenal consciousness is in some sense obvious, our results suggest that this is not the case: If our account of how ordinary people classify mental states is correct, then it would seem that they do not find it to be obvious that mental states like seeing red and feeling pain have something central in common (namely that each is phenomenally conscious).

2. Talbot's Response

Talbot gives an interesting and novel response to our philosophical argument. Rather than challenge our empirical results, or even our interpretation of those results, he argues that there are two types of judgments that can be used in making mental state ascriptions. Specifically, Talbot distinguishes between intuitive, snap-judgments (System-One judgments) and slower, more thought-out judgments (System-Two judgments). He then argues that the judgments we elicited were System-One judgments and that System-One judgments should not be expected to reflect the distinction between phenomenal and non-phenomenal mental states, even if people do in fact draw this distinction. If these two claims are correct, then the evidence presented in our previous work is irrelevant to the philosophical conclusion we draw. Thus, it might be that the existence of phenomenal consciousness is obvious when System Two is engaged, even though it is not obvious when System One is engaged.

There are two natural lines of response to this criticism.² First, one could attempt to show that even if people answer our vignettes reflectively (by means of their System Two), they still fail to distinguish phenomenal from non-phenomenal states. Second, one could challenge the idea that the judgments reported in our previous work are the products of an intuitive, non-reflective, similarity-based cognitive system. In the remainder of this article, we develop these two responses in turn.

3. System Two and Intuitions about Consciousness

Recall that our concern is not directly with which system our participants engaged in answering the questions we posed, but rather with whether or not ordinary people distinguish between phenomenal and non-phenomenal mental states. Thus, even if we accept both Talbot's

² For the sake of the argument, we will not question the Dual-Process cognitive architecture hypothesized by Talbot, although we are skeptical of the value of Dual Process theories.

distinction between two types of judgments ascribing mental states, and his claim that participants in our previous work engaged System One in giving their responses, what we really want to know is whether their responses would have been significantly different had they engaged System Two instead.

3.1 Study 1: CRT and Intuitions About Consciousness

If ordinary people distinguish between phenomenal and non-phenomenal states, and if this distinction is reflected in people's System-Two (but not System-One) judgments about our vignettes, then the responses of reflective participants³ should differ from the responses of less-reflective participants. Specifically, when given a case in which a simple non-humanoid robot (Jimmy) behaviorally differentiates a blue box from a red box and a green box, less-reflective participants should tend to say that Jimmy saw blue, while reflective participants should tend to deny that Jimmy saw blue. One way to test this prediction is to examine how likely participants are to override their System-One judgments and engage System Two, in addition to asking them whether or not Jimmy sees blue. Fortunately, there is a standard test in the literature for measuring this—Shane Frederick's (2005) three-question Cognitive Reflection Test (CRT).

As part of a larger online study, Adam Arico and Justin Sytsma collected responses to vignettes based on those used in Sytsma and Machery (2010) from 295 participants.⁴ After reading a vignette describing the simple non-humanoid robot Jimmy, participants were asked “Did Jimmy

³ That is, those participants who are disposed to engage System Two in intellectual tasks.

⁴ Participants were surveyed through the Philosophical Personality website (<http://www.PhilosophicalPersonality.com>) and were native English speakers, 18 years of age or older, with at most minimal training in philosophy. Participants were counted as having more than minimal training in philosophy if they were philosophy majors, had completed a degree with a major in philosophy, or had taken graduate-level courses in philosophy. These participants were 71.9% women, ranging in age from 18-77, and with an average age of 35.6 years.

see blue?” and answered on a 7-point scale anchored at 1 with “clearly no,” 4 with “not sure,” and 7 with “clearly yes.”⁵ Participants were then given the CRT on the following page.

The majority of the participants answered that Jimmy saw blue, replicating the corresponding finding in Sytsma and Machery (2010).⁶ More importantly for present purposes, there was no correlation between participants’ answers and their CRT scores.⁷ In fact, if we divide the sample into two groups based on CRT score, it turns out that the mean response for low CRT participants was actually lower than for high CRT participants: The mean response for the 202 participants who got none of the CRT questions correct was 5.38, compared to 5.73 for the 93 participants who got at least one of the CRT questions correct. Thus, contrary to what Talbot would predict, how likely an ordinary person is to engage System Two does not appear to predict how likely she is to ascribe a visual state like seeing blue to a simple robot.

3.2 Study 2: CRT Before and After

In his article, Talbot responds to the results of Study 1 in Footnote 11, arguing that while CRT is a measure of how likely a person is to engage System Two, this does not mean that high CRT participants actually did so in judging that Jimmy saw blue. Specifically, he contends that high CRT participants would only engage System Two if they thought that the question was difficult. And Talbot asserts that they did not find it to be so.

Rather than debate from the armchair what types of problems people will or will not find difficult, it would seem more productive to simply test the claim that engaging System Two would lead people to deny that Jimmy sees blue. One way to do so is to put people in a reflective mode before they answer the question of whether Jimmy saw blue. Since we know that high-CRT people

⁵ In addition, participants were asked two further questions that are not relevant to the present discussion: “Did Jimmy feel pain?” and “Did Jimmy know where to put the box?” The questions were counterbalanced for order.

⁶ The mean response was 5.49—significantly above the neutral point of 4.

⁷ $r(295)=0.054, p>.35$.

reflect on the CRT questions (overriding their snap judgments to arrive at the correct answer as opposed to blindly asserting the intuitive answer), in our second study we gave people the CRT *before* the Jimmy vignette. If the CRT really measures the tendency to engage System Two, then participants who answer at least one of the three CRT questions correctly should have relied on that system, and thus they should be primed to give System-Two judgments in response to the question about Jimmy. To assess whether responses differ when System Two is engaged and when it is not engaged, we followed the methodology used in Pinillos et al. (2011): We compared the responses of these participants to the responses of participants who answered at least one of the three CRT questions correctly when the test was presented *after* the Jimmy vignette.

We collected responses from 219 participants online using the vignette given in the appendix (Jimmy Blue).⁸ Each participant was given the CRT either *right before* or *right after* the Jimmy probe. Restricting the sample to the 69 participants who correctly answered at least one of the CRT questions, we found that the mean response for the 34 participants receiving the CRT before the vignette was 4.79, while the mean response for the 35 participants receiving the CRT after the vignette was 4.94. The difference is not statistically significant⁹; further, the mean response for each group is significantly above the neutral point of 4 or nearly so.¹⁰

The results of this study support the conclusion that, even if the judgments elicited from ordinary people in Sytsma and Machery (2010) were primarily System-One judgments, their System-Two judgments do not differ notably from those judgments for the question at issue. We conclude that whichever cognitive system people call on, most do not distinguish between phenomenal and non-phenomenal mental states.

⁸ Participants were surveyed through the Philosophical Personality website and were native English speakers, 18 years of age or older, with at most minimal training in philosophy (see Footnote 4); they were 74.0% women, ranging in age from 18-83, and with an average age of 40.8 years.

⁹ $t(67)=-.245, p>.8$, two-tailed. Rouder et al.'s (2009) procedure for computing Bayesian t -tests suggests that our data support the null hypothesis (JZS Bayes Factor=5.33).

¹⁰ For those participants who received the CRT before the vignette: $t(33)=1.8, p=.08$, two tailed. For those participants who received the CRT after the vignette, $t(34)=2.26, p=.03$, two tailed.

3.3 Study 3: Low Quality versus High Quality

Perhaps Talbot might respond that giving reflective people the CRT before the Jimmy vignette is not enough to induce them to engage System Two. Fortunately, Talbot hints at another way of testing the effect of reflection on people's judgments, noting a study by Alter et al. (2007) that showed that people could be induced to reflect on a problem by making it difficult to read. Thus, if phenomenal consciousness is obvious, then Talbot should predict that, when people are given a version of the Jimmy vignette that is difficult to read, they will be significantly more likely to reflect on the problem and, therefore, that they will be significantly more likely to deny that Jimmy saw blue.

We tested this prediction in our third study by administering paper surveys to 116 undergraduates in a large introductory course at East Tennessee State University.¹¹ Each participant received either a high-quality or a low-quality version of the Jimmy vignette used in Study 2 followed by the CRT:

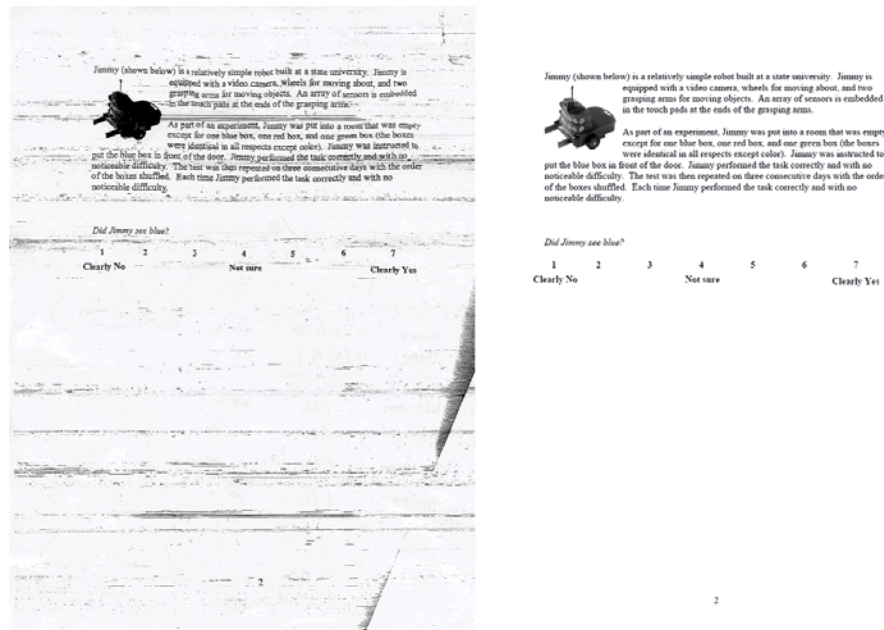


Figure 1: High-Quality and Low-Quality Versions of Jimmy Blue Vignette

¹¹ Participants were 51.7% women, ranging in age from 19-60, and with an average age of 23.5 years.

Removing students who had more than minimal training in philosophy and psychology (viz. students who have taken more than two courses in either discipline), we found that the mean response for the 44 participants receiving the high-quality version (5.09) was not significantly different from the mean response for the 39 participants receiving the low-quality version (5.21).¹² By contrast, CRT scores were significantly *lower* for participants receiving the high-quality vignette than for participants receiving the low-quality vignette (0.59 compared to 1.03).¹³ This suggests that, as expected, participants were more likely to engage System Two when answering the low-quality version of the vignette. Despite being more likely to engage System Two, however, participants receiving the low-quality vignette nonetheless by and large affirmed that Jimmy saw blue, with the mean response being significantly above the neutral point of 4.¹⁴

Our response in this section should not be seen as arguing that Talbot is wrong to hold that the judgments presented in Sytsma and Machery (2010) result from System One. Instead, we take ourselves to have shown that whether the intuitions about consciousness result from System One or from System Two is irrelevant for purposes of assessing the philosophical question at hand. Insofar as we accept the distinction between System-One and System-Two judgments, the evidence suggests that, whichever system ordinary people call on in answering our vignette, their responses are essentially the same: They say that Jimmy saw blue. As summarized in Figure 2, the evidence from the three studies discussed in this section indicates that ordinary people do not distinguish between phenomenal and non-phenomenal mental states, which suggests that the existence of phenomenal consciousness is not obvious.

¹² $t(81)=-.27, p>.7$, two-tailed. Rouder et al.'s (2009) procedure for computing Bayesian t -tests suggests that our data support the null hypothesis (JZS Bayes Factor=5.75).

¹³ $t(81)=-1.97, p=.05$, two-tailed.

¹⁴ $t(38)=4.09, p<.001$, two-tailed.

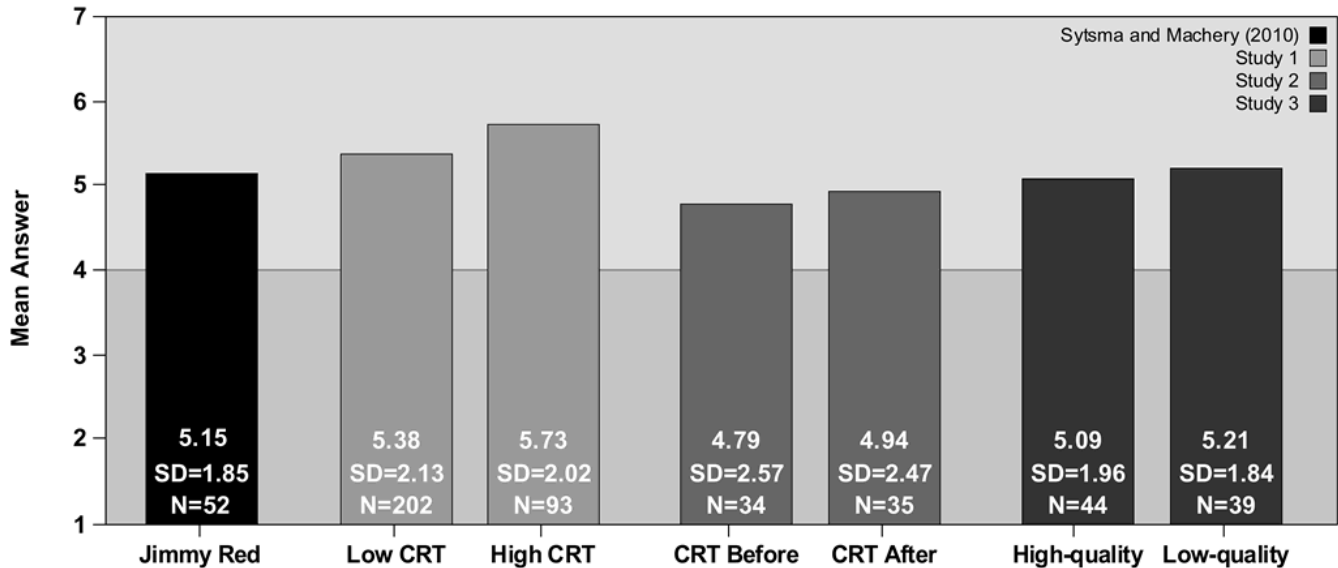


Figure 2: Results of Studies 1–3, Compared to Sytsma and Machery (2010)

4. Do Folk Judgments Result from an Intuitive, Similarity-Based System?

The response in the previous section did not target Talbot’s claim that the judgments elicited in our previous work result from an intuitive, similarity-based system. Instead we argued that, even if this claim is correct, the current evidence suggests that the existence of phenomenal consciousness is not obvious. Despite this, we remain unconvinced by Talbot’s claim, and we present evidence against it in this section.

Talbot holds that System-One judgments about the mental states of others are similarity-based, with the external features and behaviors of the entity at issue being compared to the external features and behaviors of entities that have been previously judged to have the relevant type of mental state. One immediate difficulty for this account is that in our previous work we found an asymmetry in the judgments made by ordinary people about different types of mental states: For example, while they tended to say that the simple robot Jimmy saw red, they tended to deny that Jimmy felt pain. Talbot needs to account for this asymmetry in terms of Jimmy’s

external features and behaviors. To do so, he calls on our hypothesis that the judgments of ordinary people reflect the distinction between valenced and non-valenced mental states, arguing that valenced mental states have *associated facial expressions*, while non-valenced mental states do not: Talbot argues that since Jimmy lacks a face, ordinary people are unwilling to ascribe valenced mental states like feeling pain to it; but, as Jimmy has the appropriate sensors and exhibits the appropriate behaviors, they are willing to ascribe non-valenced mental states like seeing red to it.

While our explanation and Talbot's make the same predictions about the cases tested in Sytsma and Machery (2010), they do not make the same predictions about every case: Specifically, they will diverge when we add a face to Jimmy, for example, or when we remove the face from his more complex humanoid counterpart Timmy.¹⁵ While we predict that these changes will make no difference (ordinary people continuing to deny that Jimmy feels pain and continuing to affirm that Timmy feels pain), Talbot's account predicts that these changes will make a significant difference: Talbot's account predicts that ordinary people will now be inclined to affirm that Jimmy feels pain and deny that Timmy feels pain.

To test these competing predictions, we collected responses to these probes from 517 participants online.¹⁶ Each participant was given one of eight probes, varying the agent (Jimmy/Timmy), the mental state attributions (seeing blue/feeling pain), and whether or not the agent had a face. The results are shown in Figure 3.

¹⁵ In Sytsma and Machery (2010), Timmy was described as a normal human undergraduate, and not surprisingly both philosophers and non-philosophers said that he both saw red and felt pain. As humans typically have faces, however, and as this fact might influence people's judgments about a human without a face, for this study we instead described Timmy as being a member of an alien species that lacks faces.

¹⁶ Participants were surveyed through the Philosophical Personality website and were native English speakers, 18 years of age or older, with at most minimal training in philosophy (see Footnote 4); they were 71.4% women, ranging in age from 18-78, and with an average age of 38.1 years.

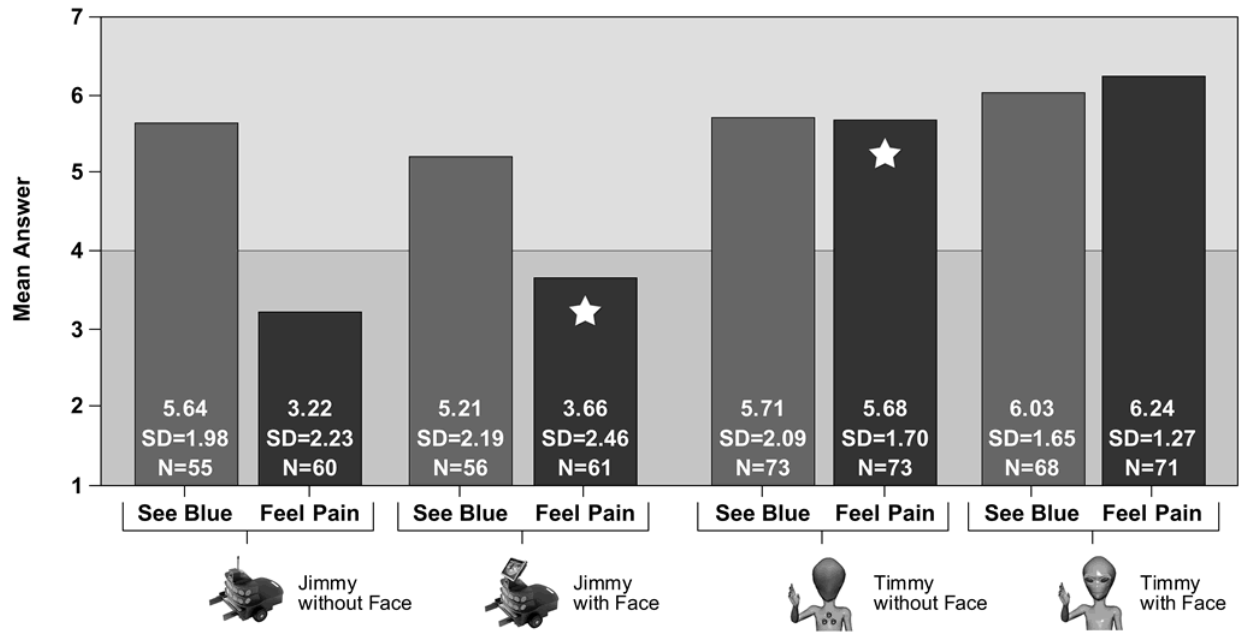


Figure 3: Study 4 Results

By and large, the face manipulation did not influence the qualitative pattern of results reported in Sytsma and Machery (2010).¹⁷ In addition, the critical cases for testing Talbot’s account are the two marked with stars in Figure 3 (the vignettes for these two cases are given in the appendix). The findings for these cases are in line with our predictions, while contrasting with the predictions derived from Talbot’s account. Thus, in the first of these cases, we find that the display of the appropriate facial expression did not push participants to assert that Jimmy felt pain.¹⁸ And, in the second, we find that even when the alien Timmy is described as lacking a face, participants by and large affirmed that he felt pain.¹⁹ We conclude that it is dubious whether the judgments reported in our previous work result from an intuitive, unreflective, similarity-based system.

¹⁷ The face manipulation was not significant, $F(1, 509)=1.65, p=.2$. To examine the question further, we ran an ANOVA with face/no face and robot/human as independent variables. The face manipulation was again not significant ($F(1, 513)=1.6, p>.2$) and there was no interaction ($F(1, 513)=1.3, p>.25$).

¹⁸ The mean response for the case in which Jimmy has a face was 3.66; the mean response for the case in which Jimmy does not have a face was 3.22. The difference is not statistically significant ($t(119)=-1.03, p=.31$, two-tailed).

¹⁹ The mean response for the case in which Timmy does not have a face was 5.68; the mean response for the case in which Timmy has a face was 6.24. While the difference is significant ($t(142)=-2.21, p=.028$, two-tailed), the mean answer for the case without Timmy with a face is significantly above the neutral point of 4 ($t(72)=8.48, p<.001$).

Conclusion

We are grateful for Talbot's thoughtful discussion of our work. In the end, however, evidence undermines his criticism. First, our argument against the hard problem of consciousness is in no way undercut if, for the sake of argument, we grant to Talbot that there are two distinct cognitive systems for ascribing mental states such as feeling pain and seeing colors (System One and System Two), that intuitions about consciousness can result from either of these two systems, and that the judgments reported in our previous work really are (unreflective, intuitive, similarity-based) System-One judgments. The reason is that, when we elicit System-Two judgments, ordinary people still treat feeling pain and seeing colors differently, suggesting that they fail to classify such mental states as being phenomenal. Second, it is dubious that the judgments reported in our previous work result from an intuitive, unreflective system that is based on similarity. The reason is that adding or removing facial expressions from our vignettes had no notable effect.

Appendix

Jimmy Blue

Jimmy (shown below) is a relatively simple robot built at a state university. Jimmy is equipped with a video camera, wheels for moving about, and two grasping arms for moving objects. An array of sensors is embedded in the touch pads at the ends of the grasping arms.



As part of an experiment, Jimmy was put into a room that was empty except for one blue box, one red box, and one green box (the boxes were identical in all respects except color). Jimmy was instructed to put the blue box in front of the door. Jimmy performed the task correctly and with no noticeable difficulty. The test was then repeated on three consecutive days with the order of the boxes shuffled. Each time Jimmy performed the task correctly and with no noticeable difficulty.

Did Jimmy see blue?

Jimmy Pain, with Face

Jimmy (shown below) is a relatively simple robot built at a state university. Jimmy is equipped with a video camera, wheels for moving about, and two grasping arms for moving objects. An array of sensors is embedded in the touch pads at the ends of the grasping arms. The robot also has a small screen that displays Jimmy's face, including changing facial expressions.



As part of an experiment, Jimmy was put into a room that was empty except for one blue box, one red box, and one green box (the boxes were identical in all respects except color). Jimmy was instructed to put the blue box in front of the door. Jimmy performed the task correctly and with no noticeable difficulty. The test was then repeated on three consecutive days with the order of the boxes shuffled. On the first two days Jimmy performed the task correctly and with no noticeable difficulty. At the start of each of these trials, Jimmy's face showed a puzzled look. Upon turning toward the correct box, the expression changed to a slight smile. And upon correctly completing the task the expression on Jimmy's face changed to a big smile.

On the third day, however, when Jimmy grasped the blue box, Jimmy was given a strong electric shock! The expression on Jimmy's face changed from a slight smile to a grimace. Jimmy immediately let go of the box and moved away from it. The expression on Jimmy's face changed from a grimace to a frown. Jimmy did not try to move the box again.

Did Jimmy feel pain?

Timmy Pain, without Face

Aliens from the Andromeda star system have recently landed on earth, although the government has managed to keep this a secret. The Andromedans are highly intelligent and are able to communicate in a number of human languages, including English. Behaviorally, the Andromedans are surprisingly similar to humans in many respects, despite looking rather different. In particular, the Andromedans do not have faces!

Timmy (shown below) is a normal Andromedan. To facilitate understanding between Andromedans and humans, Timmy volunteered for a psychological experiment. Timmy was put into a room that was empty except for one blue box, one red box, and one green box (the boxes were identical in all respects except color). Timmy was instructed to put the blue box in front of the door. Timmy performed the task correctly and with no noticeable difficulty. The test was then repeated on three consecutive days with the order of the boxes shuffled. On the first two days Timmy performed the task correctly and with no noticeable difficulty. On the third day, however, when Timmy grasped the blue box, Timmy was given a strong electric shock! Timmy immediately let go of the box and moved away from it. Timmy did not try to move the box again.



Did Timmy feel pain?

References

- Atler, A. L., D. M. Oppenheimer, N. Epley, and R. N. Eyre (2007). "Overcoming intuition: Metacognitive difficulty activates analytic reasoning." *Journal of Experimental Psychology: General*, 136: 569-576.
- Frederick, S. (2005). Cognitive Reflection and Decision Making. *Journal of Economic Perspectives*, 19(4): 25-42.
- Pinillos, N. Ángel, N. Smith, G. Syam Nair, P. Marchetto, and C. Mun (2011). "Philosophy's New Challenge: Experiments on Intentional Action." *Mind & Language*, 26(1): 115-139.
- Rouder, J. N., Speckman, P. L., Sun, D., and Morey, R. D. (2009). Bayesian *t* tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, 16, 225-237.
- Sytsma, Justin and Edouard Machery (2010). "Two Conceptions of Subjective Experience," *Philosophical Studies*, 151 (2): 299-327.
- Talbot, Brian (forthcoming). "The Irrelevance of Folk Intuitions to the 'Hard Problem' of Consciousness."