EVIL INTUITIONS? The Problem of Evil and Experimental Philosophy

Ian Church, Isaac Warchol, Justin Barrett

ABSTRACT

In this paper, we apply the resources of experimental philosophy to William Rowe's seminal version of the problem of evil, as it is found in his 1979 article, "The Problem of Evil and Some Varieties of Atheism." We argue that a key intuition that is driving Rowe's formulation of the problem is not shared across various demographics. We argue that such a result raises serious philosophical questions regarding the ultimate success of Rowe's argument.

INTRODUCTION

In William Rowe's seminal version of the problem of evil, he levels the following argument against theism:

1. There exist instances of intense suffering which an omnipotent, omniscient being could have prevented without thereby losing some greater good or permitting some evil equally bad or worse.
2. An omniscient, wholly good being would prevent the occurrence of any intense suffering it could, unless it could not do so without thereby losing some greater good or permitting some evil equally bad or worse.
3. [Therefore,] there does not exist an omnipotent, omniscient, wholly good being. (1979, 336)

Like Rowe, let's use the following shorthand when discussing this argument: an instance of suffering is *pointless* if allowing it to happen doesn't either afford some greater good or prevent some other evil equally bad or worse. For this paper, our focus will be on premise 1.

Why should we think premise 1 is true? Here Rowe has us think about an example of what seems like a good candidate for a pointless evil:

FAWN: Suppose in some distant forest lightning strikes a dead tree, resulting in a forest fire. In the fire a fawn is trapped, horribly burned, and lies in terrible agony for several days before death relieves its suffering. (1979, 337)

According to Rowe, "so far as we can see, the fawn's intense suffering is pointless" (1979, 337). (Though, whether or not the suffering is *actually* pointless is the subject of the debate.) While an omnipotent, omniscient, all-good being certainly could have prevented such an event, it's extremely difficult to imagine how permitting something like the suffering of FAWN could either prevent a greater evil from occurring or might usher in some greater good. As such, premise 1 looks plausible.

But, as Rowe is quick to note, this doesn't amount to a proof. For all we can tell, there *is* a greater evil that allowing FAWN prevents or perhaps there *is* a greater good that allowing FAWN affords. The problem, as Rowe sees it, is that given "*our* experience and knowledge of the variety and profusion of suffering in our world" it sure seems like evils like those manifest in FAWN are wholly avoidable and more or less pointless; and while the above argument doesn't amount to a proof, it does, according to Rowe, provide "rational support for atheism, that it is reasonable for us to believe that the theistic God does not exist" (1979, 338, emphasis ours).

Critically, it's our intuitions regarding FAWN that are the driving force for thinking that premise 1 is true. As such, if we don't think that the suffering of the FAWN is pointless, contrary to Rowe, then our evidence in favor of thinking that 1 is true greatly diminishes. And if our evidence in favor of thinking that 1 is greatly diminished, then, as Rowe rightly acknowledges, the evidence the argument generates against theism greatly diminishes too.

It's also worth noting, however, that in this seminal account of the problem of evil there are important questions that can be raised from the perspective of experimental philosophy. When epistemologists talk about "our intuitions" regarding Gettier counterexamples, for example, experimental philosophers wonder, "whose intuitions?" "Does everyone have such intuitions?"; similarly, when Rowe talks about "*our* experience and knowledge of the variety and profusion of suffering in our world" we might also easily wonder "whose experience?" or "whose knowledge?". Similarly, when Rowe says that it "does not appear" or doesn't "seem" reasonable to believe "that there is some greater good so intimately connected to [the suffering of FAWN] that even an omnipotent, omniscient being could not have obtained that good without permitting that suffering or some evil at least as bad", we might plausibly ask "does it 'appear' or 'seem' this way to everyone?"

ACKNOWLEDGEMENTS

HYPOTHESES

We predicted that the intuitions that underwrite Rowe's 1979 formulation of the problem would not be shared across various demographics. With this prediction, the following hypotheses emerged:

Gender: Given that men are statistically more likely to be atheists or agnostics than women (Cragun 2016:307), we predicted that men would, on average, agree with Rowe's intuition more than women.

Education: Additionally, given that education levels negatively correlate with religiosity (Beit-Hallahmi 2006:313)--such that the more educated someone is the more likely they are to be an atheist or an agnostic--we predicted that more educated people will report greater agreement with Rowe's intuition, on average, than less educated people.

Nationality: Given that Rowe is working within the American academy, we expected that Americans might, on average, be more likely to agree with Rowe than other nationalities.

Ethnicity: Given that Rowe is working within the anglophone academic world, a world that has historically been predominantly populated by people of European descent, we expected that people who identify as White would, on average, be more likely to agree with Rowe than other ethnicities.

For the purposes of this paper, these are the hypotheses we will focus on.

We also tested a range of ancillary hypotheses aimed at exploring what factors are contributing to the target philosophical intuitions (e.g. the cuteness of the animal, the inclusion of a picture of the animal, the presence of context). For the purposes of this paper, however, we won't focus on these results; we only mention these ancillary hypotheses to help explain our experiment design.

GRAPH 1

Rowe Agreement and Gender
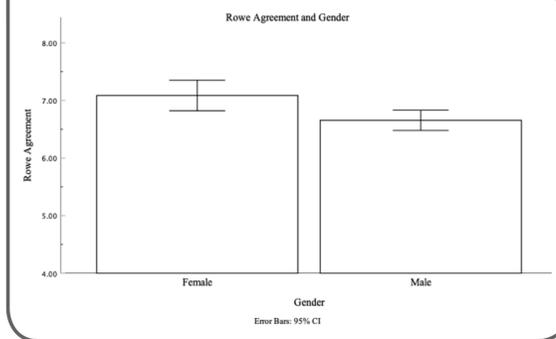
METHODOLOGY

To investigate these questions, we developed an experimental study with a 2x2x3 between-subjects factorial design. 1,506 participants were recruited from Amazon's Mechanical Turk online workforce. After completing an informed consent form, participants provided demographic information. Participants then read Rowe's vignette of the fawn from the 1979 paper. Participants were presented with the vignette in one of several manners. To half of the participants the vignette was accompanied by a description of the role of wildfires in a forest ecosystem to provide context to the suffering, the other half read the vignette without context, just as it appeared in Rowe's 1979 paper. The subject of the vignette varied as either a fawn, a boar, or a vulture. Finally, in half of the cases a picture of the subject of the vignette accompanied the vignette. Thus, this experiment contained three variables: context, picture, and animal.

After reading the vignette, participants rated several statements designed to assess their degree of agreement or disagreement with Rowe's intuition that the suffering described in the vignette is pointless. These statements read, "The story you just read is an example of pointless suffering," "Some equal or greater evil could have been prevented because of the situation in the story," and "Some equal or greater good could be accomplished because of the situation in the story." Participants responded on a 7 point Likert scale ranging from 1, Strongly Disagree to 7, strongly agree. We initially intended to measure the degree to which participants shared Rowe's intuitions through an index compiled of the score of these three statements, however, we found that whereas scores of the last two questions were highly correlated ($r = .478$, $p < 0.01$) the first question was not highly correlated in the expected direction with the last two questions ($r = .071$, $p < 0.01$ and $r = -.171$, , $p < 0.01$). Therefore, we measured agreement with Rowe through an index of the reverse scored second and third questions. Finally, the participants answered questions about their intuitions concerning pointlessness and suffering more broadly.

RESULTS

Demographics: After excluding participants who failed attention checks, rushed through the survey, or abandoned the survey, we had a sample size of $n = 1,506$. Of these 476 where female, 1,014 were male, 16 had another gender identity. The sample consisted of 846 White participants, 363 Asian participants, 146 Black or African American participants, 105 Hispanic participants, and 46 participants belonging to other ethnicities. 4 participants had a 9th grade education or less, 117 participants had a high school education or G.E.D., 158 had some college or specialized training, 82 had associates degrees, 899 had Bachelor's degrees, 246 had a Master's degree or higher.

Please note: In the following results, a score of 8 represents a midpoint of neither agreeing or disagreeing with Rowe. Anything above 8 (maximum of 14) represents agreement with Rowe on average. Anything below 8 (minimum of 2) represents disagreement with Rowe on average.

Gender: Women agreed with Rowe significantly ($p = 0.007$) more than men in an independent samples *t*-test. Both men and women showed an overall average disagreement with Rowe. When using a one sample *t*-test with a test value of 8, both men ($M = 6.66$, $SD = 2.85$, $p < 0.001$) and women ($M = 7.09$, $SD = 2.94$, $p < 0.001$) show that both groups overall significantly disagree with Rowe. [Graph 1]

Education: A Welch ANOVA revealed a significant relationship between Rowe agreement and level of education (Welch's $F(4) = 27.56$, $p < 0.001$). Opposite our hypothesis, as education level increases, agreement with Rowe decreases. Those with only a high-school education tended to agree with Rowe ($M = 8.51$, $SD = 2.76$) and those with some college or specialized training did not show a tendency to agree or disagree ($M = 8.05$, $SD = 3.10$). But those with an associate degree ($M = 7.70$, $SD = 2.96$), a bachelor degree ($M = 6.48$, $SD = 2.77$), or a postgraduate degree ($M = 6.05$, $SD = 2.57$) all showed an average tendency to disagree with Rowe. [Graph 2]

Ethnicity: The use of Welch ANOVA detected a significant relationship between ethnicity and agreement with Rowe (Welch's $F(4) = 24.64$, $p < 0.001$) . A Games-Howell post-hoc analysis revealed several significant differences between ethnic groups. The White mean ($M = 7.31$, $SD = 2.90$) was significantly higher than the Asian mean ($M = 5.89$, $SD = 2.55$, $p < .001$) as well as significantly ($p < 0.001$) higher the Black/African American mean ($M = 5.99$, $SD = 2.53$, $p < .001$). The Hispanic mean ($M = 7.17$, $SD = 3.12$) was significantly ($p = 0.002$) greater than the Asian mean and significantly higher than the Black of African American mean ($p = 0.014$). [Graph 3]

Nationality: Three nationalities were well represented in our sample, Americans, 983, Indians, 299, and Brazilians, 51. A significant Welch ANOVA (Welch's $F(2) = 60.747$, $p < 0.001$) and a Games-Howell test show significant differences between Americans ($M = 7.10$, $SD = 2.88$), Brazilians ($M = 8.47$, $SD = 2.96$) and Indians ($M = 5.41$, $SD = 2.41$). The differences between Indians and American as well as between Indians and Brazilians are significant (both to $p < 0.001$). The difference between Americans and Brazilians was significant to $p = 0.006$. [Graph 4]

Experimental Results: Participants were exposed to the FAWN vignette in different forms, of which the high and low context groups yielded highly significant results. To ensure that the above demographic findings were not attributable to the effects of context, multiple univariate analyses were conducted with demographic variables as covariates. While statistically controlling for context, ethnicity, religion, nationality, and education remained significant to the $p < 0.001$ degree and gender remained significant ($p = 0.022$).

The team planned to use a one way ANOVA and a Tukey Honestly Significant Difference post-hoc test, however, Levene's test showed that the assumption of homogeneity had been violated so a Welch Test and a Games-Howell test was used.
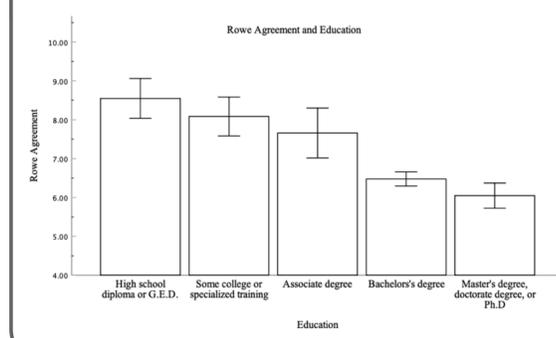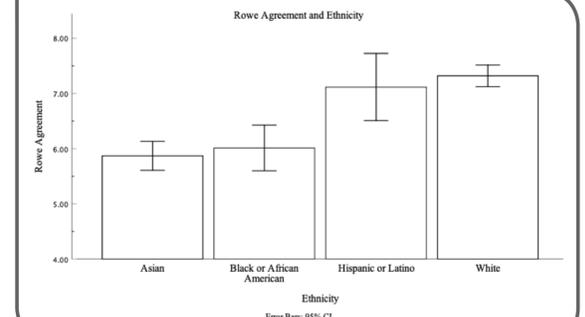
CONCLUSION

Across those five demographic variables - gender, education, ethnicity, religion, and nationality - we saw significant divergences from Rowe's intuition. While intuitions do seem to diverge according to gender, it's women (not men) who are more likely to agree with Rowe regarding the FAWN case. And while intuitions also seem to diverge according to education, it was the least educated (not the most educated) who were most likely to agree with Rowe.

But perhaps what is most striking from all of this research is the fact that so few people across all of these demographics agree, on average, with Rowe's intuition regarding FAWN. While White and Hispanic participants were significantly more likely to agree with Rowe's intuition than, say, Asian participants, Whites and Hispanics nevertheless, on average, still disagreed with Rowe's intuition. While Americans and Brazilians were significantly more likely to agree with Rowe's intuition than Indians, Americans, on average, nevertheless still disagreed with Rowe's intuition regarding FAWN. While women were significantly more likely to agree with Rowe's intuition than men, women, on average, nevertheless still disagreed with Rowe's intuition.

In any case, it's now clear that the driving intuitions behind premise 1 are not shared across various demographics. But does such a variety in response to FAWN really threaten Rowe's argument? After all, a defender of Rowe's argument might argue that their conclusions about FAWN are not driven by intuition but are the result of a rational assessment of the possibility of rational justification of the target suffering! To be sure, many philosophers (e.g. Wykstra 1984; Russell and Wykstra 1988; Inwagen 1988; Alston 1991; Plantinga 2000) have already cast doubt on such a response; however, given the details of our empirical research, such a response now might seem especially implausible. Do we have a good reason for thinking that as people become more educated, they become generally less able to rationally assess FAWN? That people of European decent are better at rationally assessing FAWN than Blacks or Asians? Surely not. As such, our empirical findings raise serious concerns about the ultimate success of Rowe's seminal formulation of the problem of evil, since it seems to suggest that Rowe's response to FAWN might be underwritten by cognitive mechanisms/influences that are not nearly as reliable, universal, or objective as we might have hoped.

GRAPH 3

Rowe Agreement and Ethnicity

GRAPH 2

Rowe Agreement and Education

GRAPH 4

Rowe Agreement and Nationality