EMPIRICAL CHALLENGES TO THE EVIDENTIAL PROBLEM OF EVIL

BY BLAKE MCALLISTER, IAN CHURCH, PAUL REZKALLA, AND LONG NGUYEN

Forthcoming in *The Oxford Studies in Experimental Philosophy* series

The problem of evil is broadly considered to be one of *the* greatest intellectual threats to traditional brands of theism.[1] And William Rowe's 1979 formulation of the problem in "The Problem of Evil and Some Varieties of Atheism" is the most cited formulation in the contemporary philosophical literature. In this paper, we explore how the tools and resources of experimental philosophy might be brought to bear on Rowe's seminal formulation, arguing that our empirical findings raise significant questions regarding the ultimate success of Rowe's argument. Such a result would be quite notable within philosophy of religion, since this is considered one of the most formidable arguments against theism. However, further testing is needed before any firm conclusions can be drawn.

In section 1, we elucidate Rowe's formulation of the problem of evil and the intuitions that seem to underwrite it. In section 2, we explore how the tools and resources of experimental philosophy might be brought to bear on Rowe's formulation, outlining our hypotheses and our methods for testing them before showcasing our results. In section 3, we discuss the philosophical import of our results–arguing that our results, when taken together, pose an initial challenge to Rowe's seminal argument.

1. Rowe's Problem of Evil

The problem of evil encompasses a family of arguments, all of which contest theism on the basis of some supposed tension between a good God and observed evils. Notable among this family of arguments is Rowe's evidential problem of evil (Rowe 1979; 1991; 2006). Rowe's argument proceeds on the incompatibility of God and gratuitous (or pointless) evils, defined as evils that an omnipotent being "could have prevented without thereby losing some greater good or permitting some evil equally bad or worse" (Rowe 1979, 336). Put succinctly:

1. If God exists, then there are no gratuitous evils.
2. There are gratuitous evils.
3. So, God does not exist.

If the incompatibility of God and gratuitous evils is granted (though perhaps it shouldn't be; see Hasker 1992, Peterson 1982, van Inwagen 1988; 1991; 2000), then the plausibility of the argument hinges on its second premise: that there are in fact gratuitous evils.

The name "the evidential problem of evil" arises out of Rowe's attempt to build an inductive case for the existence of such evils. Essentially, for certain instances or kinds of evil, we aren't aware of any reasons (greater goods or worse evils) requiring their allowance, and so likely there aren't any reasons. Rowe provides the story of a fawn trapped in a forest fire as a paradigmatic example:

---

[1] Where, for example, God is taken to be all-knowing, all-good, and all-powerful.

> 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. (Rowe 1979, 337)

Rowe goes on to reason that:

> So far as we can see, the fawn's intense suffering is pointless. For there does not appear to be any greater good such that the prevention of the fawn's suffering would require either the loss of that good or the occurrence of an evil equally bad or worse. Nor does there seem to be any equally bad or worse evil so connected to the fawn's suffering that it would have had to occur had the fawn's suffering been prevented. (Rowe 1979, 337)

In short, Rowe makes the following argument:

4. There does not appear to be any greater good or worse evil that the fawn's suffering is necessary to obtain or prevent, respectively.
5. So, there probably is no such greater good or worse evil—i.e., the fawn's suffering is gratuitous.

This style of inference has been dubbed "the noseeum inference"—roughly, I *no see* any reasons for this evil, so there must *no be* any reasons for this evil.

One way of disagreeing with Rowe would be to reject premise 4, citing some greater good or worse evil that appears connected to the fawn's suffering. This is to give a theodicy or a defense (depending on how confidently that explanation is put forward). Others have criticized the strength of the inference itself. In particular, skeptical theists have argued that our limited epistemic position with respect to these matters prevents us from concluding that the evils are gratuitous, even if no possible reasons are forthcoming (e.g., Alston 1991b, Bergmann 2001, Wykstra 1984; 1996). The prominence of skeptical theism in particular leaves Rowe's argument in something of a vulnerable position. Certainly, the position of the argument would be strengthened were there additional sources of support for premise 2 and the existence of gratuitous evils.[2]

Rowe himself hints at another potential source. Immediately following the above passages, Rowe writes, "Could an omnipotent, omniscient being have prevented the fawn's apparently pointless suffering? *The answer is obvious*" (Rowe 1979, 337; emphasis ours). In a later article, Rowe writes, "Is the fawn's suffering a pointless evil? Clearly, it certainly *seems to us* to be pointless." (Rowe 2006, 79). The suggestion perhaps is that the pointlessness of an evil might be immediately apparent rather than inferred. At the end of *Warranted Christian Belief*, the theist Alvin Plantinga reports that something like this is what he takes to be "the best version of the atheological case from evil." He writes:

> The claim is essentially that one who is properly sensitive and properly aware of the sheer horror of the evil displayed in our somber and unhappy world will *simply see* that no being of the sort God is alleged to be could possibly permit it. (Plantinga 2000, 484; emphasis ours)

Plantinga's assessment resonates with that of other philosophers such as Paul Draper, who proposes that poignant experiences of evil might immediately prompt the conviction that God would never

---

[2] And given the empirical nature of premise 2, it makes sense to wonder if these additional resources might be found via empirical research.

allow it (Draper 1991). We think it plausible that most people who find Rowe's argument compelling do so because of intuitions like these. Evidence indicates that our moral conclusions are, in most cases, predominantly shaped non-inferentially by our emotional reactions rather than by careful reasoning (e.g., Kelly 2011, Haidt 2012). Thus, we think it likely that the conclusion that some evil is gratuitous, and so morally unallowable, usually proceeds from intuition rather than a careful survey of possible reasons.[3]

Even so, does the mere intuition that an evil is pointless provide rational support for that conclusion? According to the common sense tradition, possibly yes. The common sense tradition in epistemology arguably begins as far back as Aristotle (Shields 2013), but is most strongly associated with Thomas Reid and Scottish common sense philosophy. Certain elements of Reid's epistemology, including his common sense response to skepticism, were preserved in G.E. Moore (1939; 1959) and reemerged on the contemporary scene through the work of people like Roderick Chisholm (1977), Alvin Plantinga and Nicholas Wolterstorff (1983), William Alston (1991a), and Richard Swinburne (2001; 2004). Since then, the common sense tradition has gained increasing prominence in the form of phenomenal conservatism, first articulated by Michael Huemer (2001), and perceptual dogmatism, as defended by James Pryor (2001) and others. Though the particulars vary, the common thread among such views is that non-inferentially justified beliefs are not limited to those that are absolutely certain or infallible or incorrigible, contra classical foundationalism. For instance, Huemer's phenomenal conservatism (Huemer 2007, 30) says:

> PC   If it seems to S that p, then, in the absence of defeaters, S thereby has at least some
> degree of justification for believing that p.

On this principle, *anything* that seems true might be non-inferentially justified, so long as there are no additional considerations that rebut the content of that appearance or undermine its reliability. While other common sense principles are more restrictive, they all open the doors to a broader variety of non-inferentially justified beliefs.

Indeed, some open the door so wide that belief in the existence of gratuitous evils is a candidate for non-inferential justification (Dougherty 2008; Gellman 2014). This is certainly true on Huemer's phenomenal conservatism and perhaps also perceptual dogmatism, assuming that it is possible for gratuitousness to be embedded in the content of a perceptual experience (as in Gellman 2017). So long as it seems that certain evils are gratuitous, or one experiences them as such, then one has some initial justification for believing that they are gratuitous. Other iterations of common sense epistemology may allow for this as well if, for instance, one immediately believes that those evils are gratuitous via properly functioning faculties. For convenience, let's say that one has "the intuition" that some evil is gratuitous when one has a seeming, a perceptual experience, a snap judgment, etc., to the effect that the evil is gratuitous—the sort of state that, given common sense epistemic principles, might grant *prima facie* non-inferential justification. Plausibly, intuitions of gratuitousness or pointlessness grant immediate rational support for the existence of gratuitous evils, strengthening Rowe's argument.[4]

---

[3] When an iteration of the problem of evil is supported through intuition rather than inference, it is called a "common sense" version of that argument and is an instance of the common sense problem of evil. It receives this name because proponents appeal to principles within the common sense epistemic tradition to substantiate their appeal to intuition (as discussed below).

[4] Whether intuitions of gratuitousness really can immediately justify belief in gratuitous evils is, of course, a matter of some controversy. Dissenters include Bergmann 2012, Senor 2014, and Tweedt 2015.

That being said, the justification provided by such intuitions is by no means indefeasible. If there are reasons to doubt or even seriously bring into question the reliability of those intuitions, then the rational support they provide can be defeated. For instance, several skeptical theists have appealed to our impoverished epistemic positions vis-à-vis gratuitousness as a defeater for this non-inferential justification (Matheson 2011, Rutledge 2019, and Hendricks n.d.). Note that defeat is sometimes partial. For example, the rational support provided by an especially strong intuition of gratuitousness may be diminished by skeptical theist considerations without being altogether eliminated (Dougherty 2011).

Here's what all this amounts to. The success of Rowe's evidential problem of evil can be measured by how many people receive rational support for atheism on the basis of this argument, and how strong that support turns out to be. The most contentious premise in the argument is premise 2—that gratuitous evils exist. Rowe's fawn is put forward as a prime example. Maybe some infer that this evil is gratuitous after surveying possible reasons for allowing it, but we expect that people usually intuit its gratuitousness immediately. In either case, the ultimate success of Rowe's argument will depend on *how many* agree with Rowe that there are gratuitous evils, *how strongly* they agree that there are gratuitous evils, and whether there are any *defeaters* that bring this agreement (or the intuitions underlying it) into question.

Conveniently, all of these states of affairs can be probed through experimentation (as first noted in Church, Carlson, and Barrett 2020). Using Rowe's chosen example, Ian Church and Justin Barrett designed a study to gauge how many agree (and how strongly) that the suffering in cases like that of Rowe's fawn (i.e. fawn-style cases or fawn cases) is gratuitous. After describing this study and its results, we discuss their significance. We argue that our findings raise serious questions about the overall success of Rowe's argument.

2. Testing Rowe's Argument

2.1. Methods and Hypotheses

To start, Church, Warchol, and Barrett (hereafter Church et al.) predicted that intuitions regarding fawn-style cases would indeed diverge across various demographics. Quoting from their submitted manuscript, they came up with the following hypotheses about the demographics of the intuitions in these cases:

> *Religion*: Intuitions regarding Rowe's case will significantly diverge according to the respondents' religious beliefs. More specifically, people who report being atheists or agnostics will, on average, agree with Rowe's intuitions regarding the FAWN case; whereas, people who are not atheists or agnostics will, on average, disagree with Rowe's intuitions regarding the FAWN case.

> *Gender*: Relatedly, given that men are statistically more likely to be atheists or agnostics than women (Cragun 2016, p. 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, p. 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. (Church, Warchol, and Barrett, unpublished manuscript)

Let's call these the *religion-hypothesis*, *gender-hypothesis*, and so on, respectively.

They also inquired into the psychological mechanisms that underwrite intuitions regarding fawn cases. In addition to directly testing the stability of key philosophical intuitions across various demographics, it is also worth exploring what factors might contribute to people having the intuitions that they do. We might think of this as the *psychology of philosophy* or the *psychology of philosophers* (heaven help us!).

For example, given that fawns are exceptionally cute animals (just think of Bambie!), we might wonder if cuteness is a driving factor behind our intuitions regarding the fawn case. Would horrific death of a less cute animal—for example, perhaps a boar or a vulture—be seen as any less pointless? And what if we tried to bring the cuteness to the fore by presenting people with the picture of a cute fawn along with Rowe's original case? Or if they're reading a version of the vignette with a boar or vulture instead, what if we included a less than flattering picture of the boar or vulture respectively? Call the idea that cuteness is driving the perception of pointlessness the *cuteness-hypothesis*.

We might also wonder if the brevity of the fawn case—being only two sentences long—is contributing to the perception of pointlessness. Many scholars have highlighted the importance of context and narrative when wrestling with the problem of suffering,[5] so we might wonder if the presence of background information might diminish the perception of pointlessness. Call this the *context-hypothesis*.

Additionally, we can ask whether an individual's exposure to animal death, or the lack thereof, affects how they perceive the pointlessness of the fawn's death. For example, butchers and hunters are regularly exposed to animal death. It's a part of the job description! So, it may be that for these individuals, the tendency to see the fawn's death as pointless may not be as consistently evidenced. If so, then this might suggest that for those of us who are not butchers or hunters, our intuition about the pointlessness of the fawn's death has something to do with the lack of regular exposure to animal death. In any case, Church et al. predicted that the more exposure someone has to the hunting or butchering of animals, the less likely they are to agree with Rowe's intuitions regarding fawn-style cases. Call this the *hunter/butcher-hypothesis*.

To investigate these questions, Church et al. 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 including: age, gender, ethnicity, religious affiliation, nationality, income, and education level. 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. This description, approximately a paragraph in length, discussed the role occasional, small forest fires have in the health of the ecosystem by clearing away dead organic material and helping

---

[5] Eleonore Stump's 2012 work, *Wandering in Darkness*, is particularly relevant here.

forest recovery by leaving behind a topsoil dense in organic materials.[6] The other half of the participants 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 (High or low) picture (Picture or no picture) and animal (fawn, boar, or vulture).

Participants were then asked to rate statements designed to measure their level of agreement or disagreement with Rowe's conclusion that the animal's suffering was pointless. Figuring out which statements to use to measure agreement with Rowe posed some challenges. Straightforwardly asking participants whether an omnipotent, wholly good being (aka "God") could have prevented the fawn's suffering without thereby losing some greater good or permitting some worse evil would face serious drawbacks. For starters, there is a significant worry that by invoking "an omnipotent, wholly good being," participants would be strongly tempted to simply parrot whatever they take to be theological (or atheological) orthodoxy. A second worry is that many participants are likely to say that God can do something even when that something is logically impossible. However, what really interests us vis-à-vis the problem of evil is whether it is *logically possible* that the fawn's suffering be removed without thereby losing some greater good or permitting some worse evil. After all, the problem of evil attempts to show that it is logically impossible for God to exist and to allow gratuitous evil, but if God can do the logically impossible, then the fact that allowing gratuitous evil is impossible for him is no hindrance to his allowing it! So unless we limit what God can do to the logically possible–as almost all philosophers and theologians throughout history have done–then Rowe's argument is immediately rendered inert. It seems, then, that the question to ask–the one that best isolates the relevant philosophical intuitions–is not whether God could remove the fawn's suffering without also removing some greater good or bringing about some worse evil, but whether it is logically possible for this to happen.

Of course, asking about "logical possibility" or "logical necessity" comes with its own challenges. Such language is far too technical to use in questions directed at a general audience. Most participants would not understand the concepts of logical possibility and necessity, and so the inclusion of this jargon would likely produce a lot of "noise" in the data.

---

[6] The high context paragraph read as follows:

> Forest fires are often viewed as some of the most dangerous and destructive natural disasters. While some fires of catastrophic size can be detrimental to forests and endanger human lives and infrastructure, smaller forest fires are actually an essential aspect of the forest ecosystem. It may seem counterintuitive that fires could be beneficial to the life of a forest, however, recent ecological research has shown that small burns play a major role in the health of an ecosystem as a whole. Fires, often resulting from lightning strikes, quickly and efficiently clear away thick undergrowth, dying trees, and the dead material that congregates on the forest floor. If left unchecked, dead organic material and undergrowth will prevent new trees and plants from taking root and being able to grow. The burnt organic material such as plants, shrubs, and animals, leave behind topsoil that is rich in nutrients from which new plant life can easily grow. Small forest fires also play an important role in preventing fires from reaching catastrophic sizes. When a fire is small, it is usually confined to burning the undergrowth and dead material on the forest floor and does not burn the tree canopy or kill the large trees of the forest. However, if a forest goes too long without a fire, the undergrowth will become so thick that when it does burn it will easily ignite not only the forest floor but also the trees themselves. Many experts attribute the record-setting fires that have been seen in recent years to decades of fire suppression in forests, which has left entire ecosystems vulnerable to catastrophic fires. Many species of plants have adapted to occasional fires and can quickly regrow burnt branches. Some trees even need fire to reproduce due to seed-cones that will only open when exposed to extreme temperature. Now, suppose in a distant forest lightning strikes a dead tree, resulting in a forest fire. In the fire a wild [Insert Animal Here] is trapped, horribly burned, lies in terrible agony for several days before death relieves its suffering.

Thus, in trying to capture the intuitions of a broad swath of the population (including people for whom English is not their first language), Church et al. went for a middle ground. They opted for everyday language that, while lacking some technical nuance, could still (hopefully) be used to triangulate onto the intuitions at work in Rowe's argument. The statements they decided upon were these: "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.

While Church et al. had good reasons for selecting these straightforward statements, it must be acknowledged that, in deviating from the conceptually rich and theism-targeted language in which Rowe's argument is originally couched, there is a significant worry that something will be lost in translation.[7] While there are legitimate worries in this vicinity, it is worth noting that they are not unique to this particular line of research. Whenever experimental philosophers survey folk intuitions on philosophically rich issues these kinds of concerns almost inevitably arise. They are part and parcel of this kind of research. That being said, we address what we take to be the most serious worry of this sort in section 3.1.

Responses were used to gauge agreement with Rowe in the following two ways. First, as a rough measure of average agreement with Rowe, Church et al. used an index of the reverse scored questions two and three. Church et al. initially intended to measure this through an index compiled of the score of these three statements, however, they 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 $r = -.171$, $p < 0.01$).[8] A score of 8 represents a midpoint of neither agreeing or disagreeing with Rowe on average. Anything above 8 (maximum of 14) represents agreement with Rowe on average. Anything below 8 (minimum of 2) represents disagreement with Rowe on average. Second, for the fawn's suffering to be gratuitous, it must be the case that it could have been prevented without losing some greater good *and* that it could have been prevented without bringing about some worse evil. Thus, to fully agree that the fawn's suffering is gratuitous, one must report agreement (a Likert score of 5 or higher) on *both* questions two and three (which ask about worse evils and greater goods, respectively). Accordingly, along with the averages, we have also divided all participants into three camps: Agreers, Agnostics, and Disagreers. The Agreers gave answers of 5 or higher on both questions. The Disagreers gave answers of 3 or less on at least one question. All others are Agnostics. The percentages of each camp will be reported alongside average agreement with Rowe.

2.2. Results

Demographics: After excluding participants who failed attention checks, rushed through the survey (finishing faster than 90 seconds), or abandoned the survey (with more than 10% of the survey incomplete), Church et al. had a sample size of *n* = 1,506. Of these 476 were 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

---

[7] Thanks to an anonymous referee for making this point. The reviewer was particularly worried about dropping any mention of "God" specifically. Though we do not find this particular worry particularly worrying (we think logical possibility is the more relevant target for the reasons given above), the referee's comments prompted for us a second, more formidable objection that we address at length in section 3.1.

[8] Given that Rowe uses "pointlessness" as a shorthand for not bringing about a greater good or preventing a greater evil, it makes sense to prioritize the second and third items on this index, given that the second and third items correlate with each other but not with the first item; however, that said, it's worth noting that using the three item index or the just first item would not radically change the results.

participants belonging to other ethnicities. 201 participants were agnostic, 161 atheist, 464 Catholic, 261 Hindu, 181 Protestant, 100 were another denomination of Christian, and 138 participants reported another religious affiliation. 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.

Gender: Women agreed with Rowe significantly more than men [t (1488) = -2.690, p= 0.0007, d=0.15]. 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. [Figure 1]
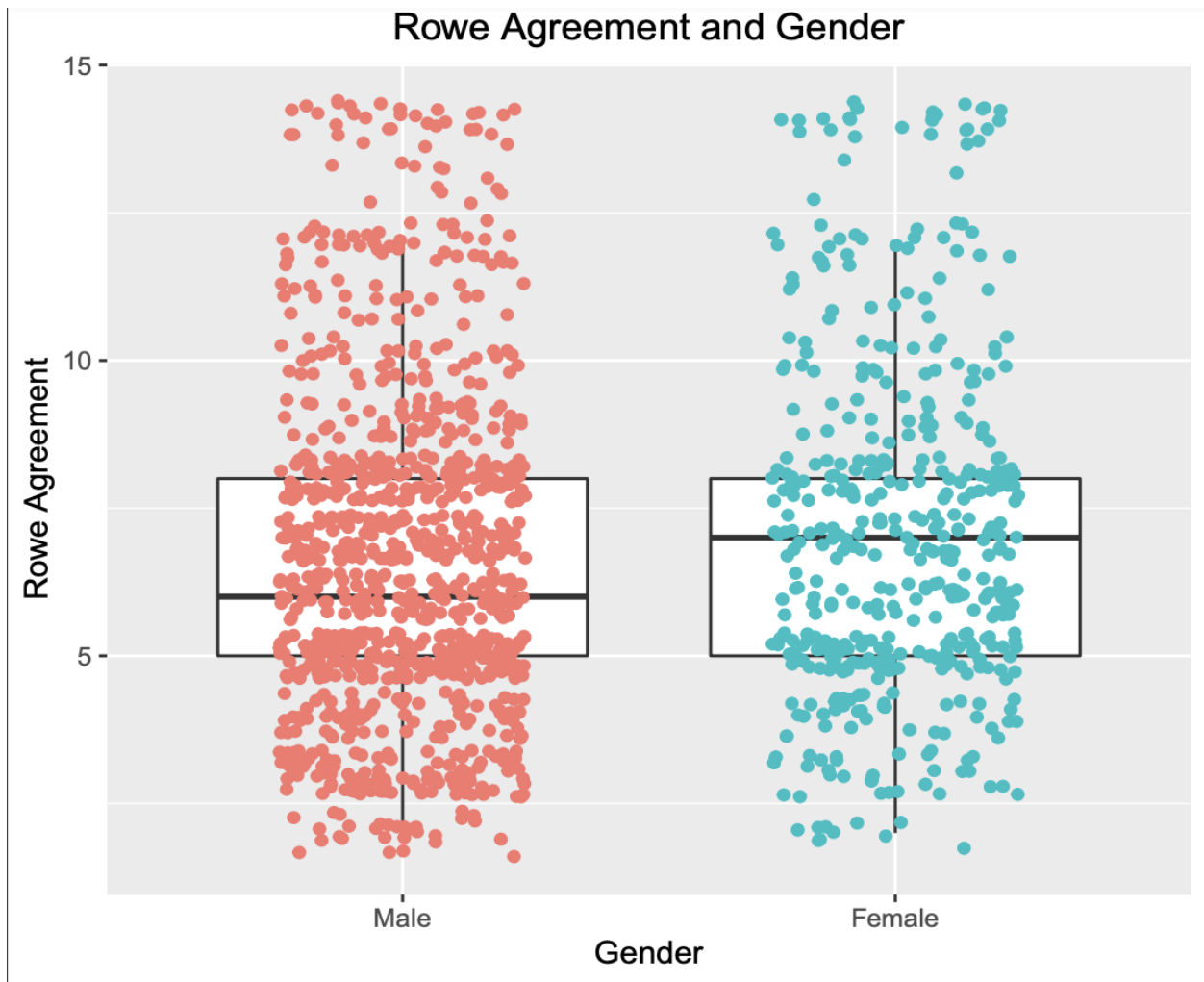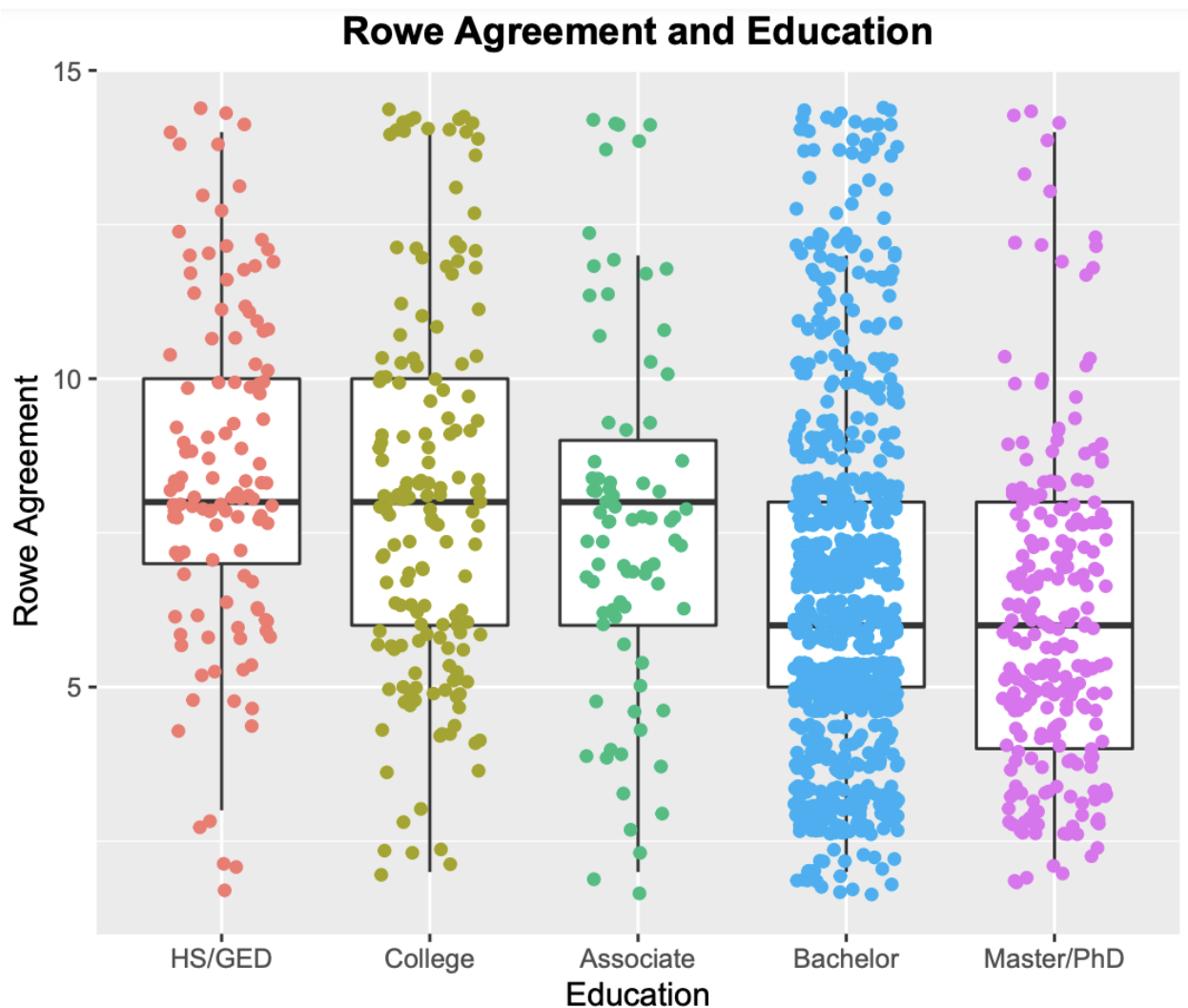

Figure 1. Rowe Agreement and Gender

For men and women the breakdown between Agreers, Disagreers, and Agnostics was as follows:

Gender

	Agreers	Agnostics	Disagreers

Male	11.6%	11.6%	76.7%
Female	14.5%	12.2%	73.3%

Education: A Welch ANOVA[9] revealed a significant relationship between Rowe agreement and level of education (*Welch's* $F(4) = 27.56$, $p < 0.001$, $\omega^2 = 0.07$). Opposite Church et al.'s 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. [Figure 2]

Rowe Agreement and Education

---

[9] 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.

Figure 2. Rowe Agreement and Education

With respect to education, the breakdown between Agreers, Disagreers, and Agnostics was as follows:

Education

	Agreers	Agnostics	Disagreers
Masters/Doc/PhD	6.5%	9.6%	83.7%
Bachelor	10.8%	9.3%	79.9%
Associate	18.3%	12.4%	68.3%
Some College/Training	22.7%	17.7%	59.5%
High School/GED	21.4%	27.4%	51.3%

Ethnicity: The use of Welch ANOVA detected a significant relationship between ethnicity and agreement with Rowe (*Welch's* $F(4) = 24.64$, $p < 0.001$, $\omega^2 = 0.07$). 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 or African American mean ($p = 0.014$). [Figure 3]
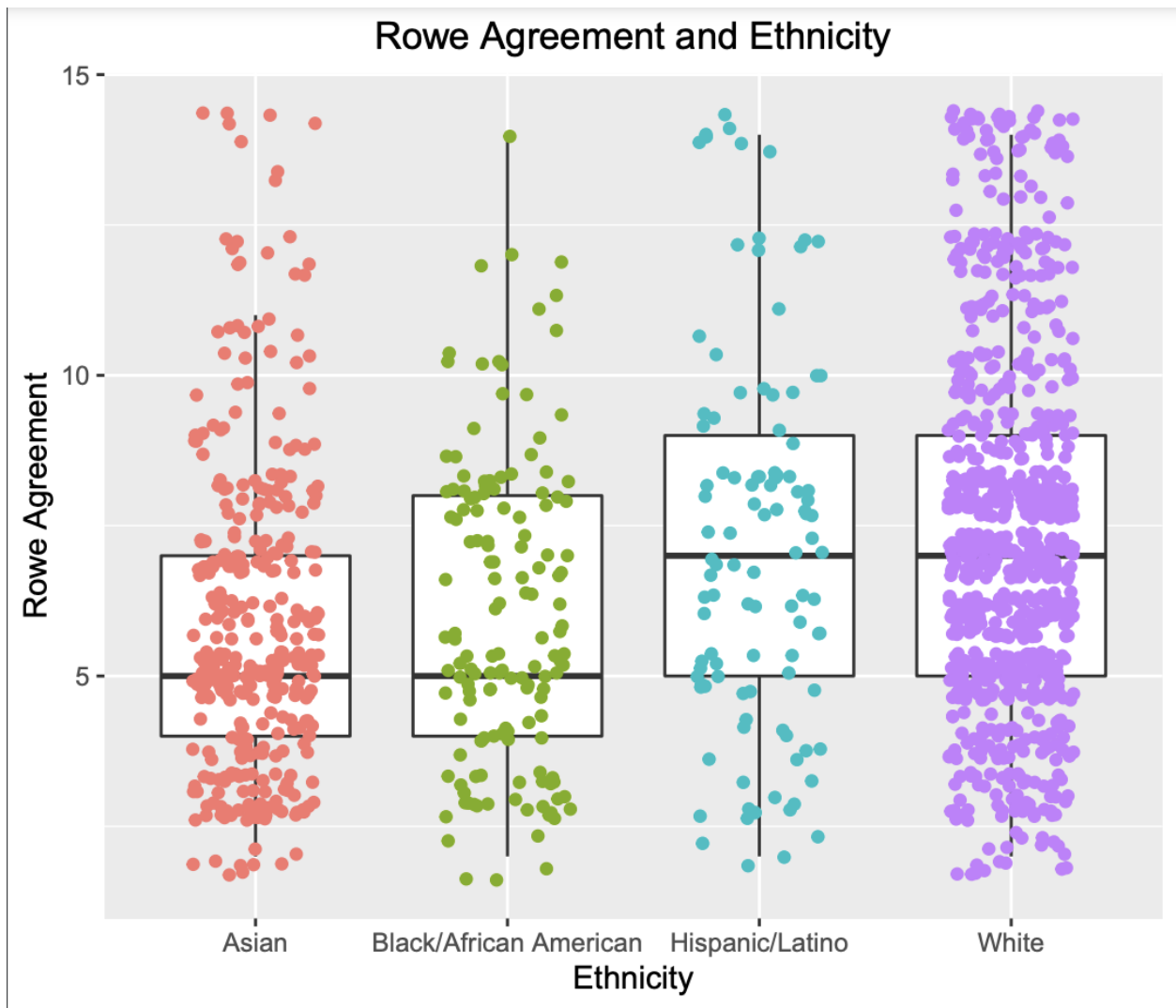
Figure 3. Rowe Agreement and Ethnicity

With respect to ethnicity, the breakdown between Agreers, Disagreers, and Agnostics was as follows:

Ethnicity

	Agreers	Agnostics	Disagreers
Asian	7.7%	8.0%	84.3%
Black/African Amer	5.5%	13.7%	80.8%
Hispanic/Latino	15.2%	17.1%	67.6%
White	15.6%	12.8%	71.6%

Religion: Another Welch ANOVA showed a significant ($p < 0.001$) relationship between agreement with Rowe and religious affiliation (*Welch's* $F(5) = 46.52$, $p < 0.001$, $\omega^2 = 0.05$). A Games-Howell post-hoc test shows the significant differences between religious groups. Both Agnostics ($M = 8.22$, SD = 2.92) and Atheists ($M = 8.58$, SD = 3.20) score significantly higher than Catholics ($M = 6.10$, SD = 2.50 $p < 0.001$) and Hindus ($M = 5.32$, SD = 2.32, $p < 0.001$) and participants claiming "Other Christian" affiliations ($M = 6.53$, $SD = 2.49$, $p < 0.001$). Atheists agree with Rowe more than Protestants ($M = 7.29$, SD = 2.67 $p = 0.001$), additionally Agnostics also agree with Rowe more than Protestants, although to a less significant degree ($p = 0.015$). [Figure 4]
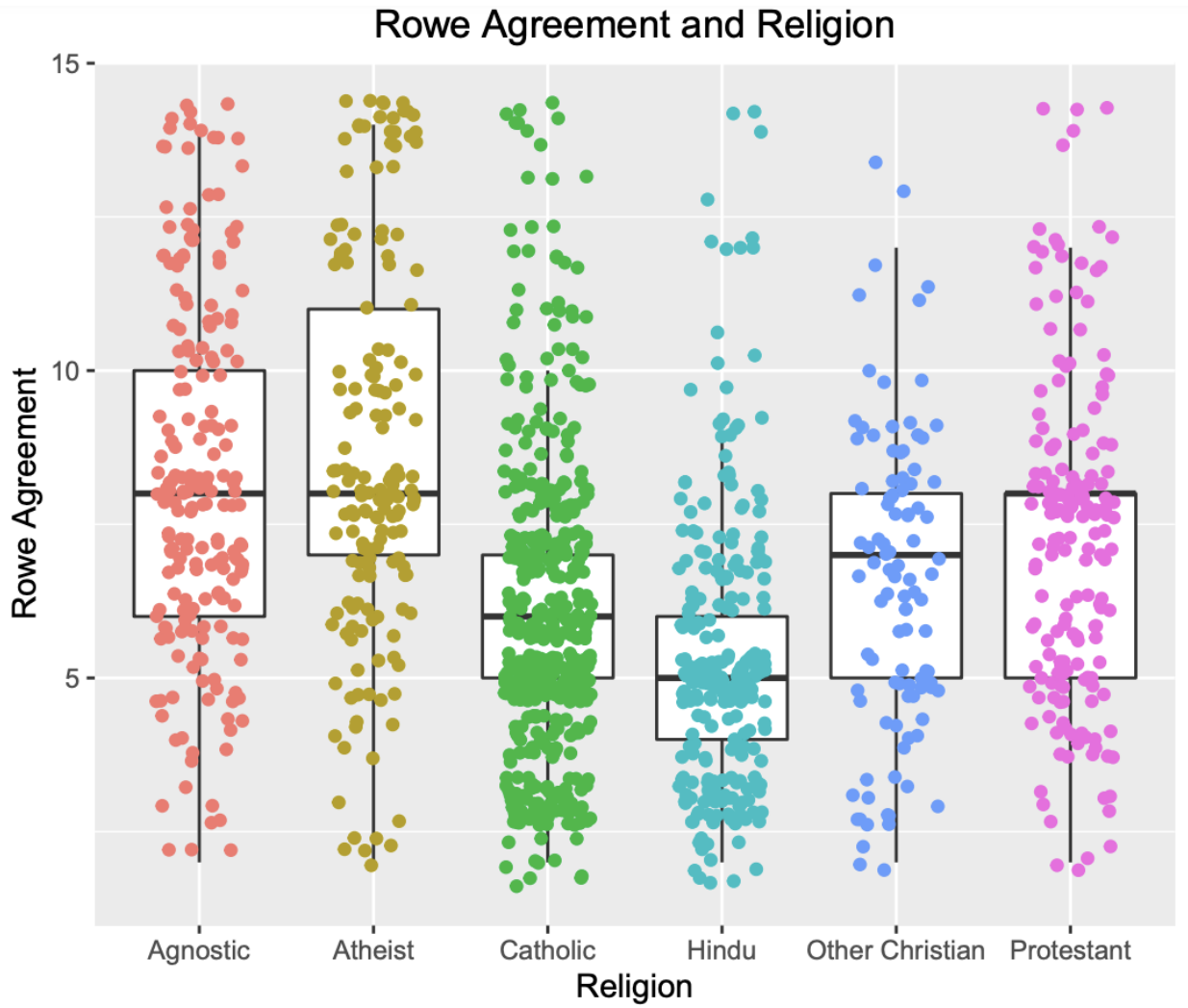

Figure 4. Rowe Agreement and Religion

With respect to religion, the breakdown between Agreers, Disagreers, and Agnostics was as follows:

	Agreers	Agnostics	Disagreers
Atheists	29.8%	14.9%	55.3%
Agnostics	24.4%	17.4%	58.2%
Catholic	5.8%	10.1%	84.1%
Protestant	15.5%	13.8%	70.7%
Other Christian	4.9%	14.9	80.2%
Hindu	0.8%	4.7%	94.5%

Nationality: Three nationalities were well represented in our sample: Americans (983) Indians (299) and Brazilians (51)— participants with other nationalities were not sufficiently numerous for Church et al. to say anything of statistical significance about them. A significant Welch ANOVA (*Welch's F*(2) = 60.747, $p < 0.001$, $\omega^2 = 0.07$) and a Games-Howell test show significant differences between Americans (*M* = 7.10, SD = 2.88), Brazillians (*M* = 8.47, SD = 2.96) and Indians (*M* = 5.41, SD = 2.41). The differences between Indians and Americans as well as between Indians and Brazilians are significant (both to $p < 0.001$). The difference between Americans and Brazillians was significant to $p = 0.006$. [Figure 5]
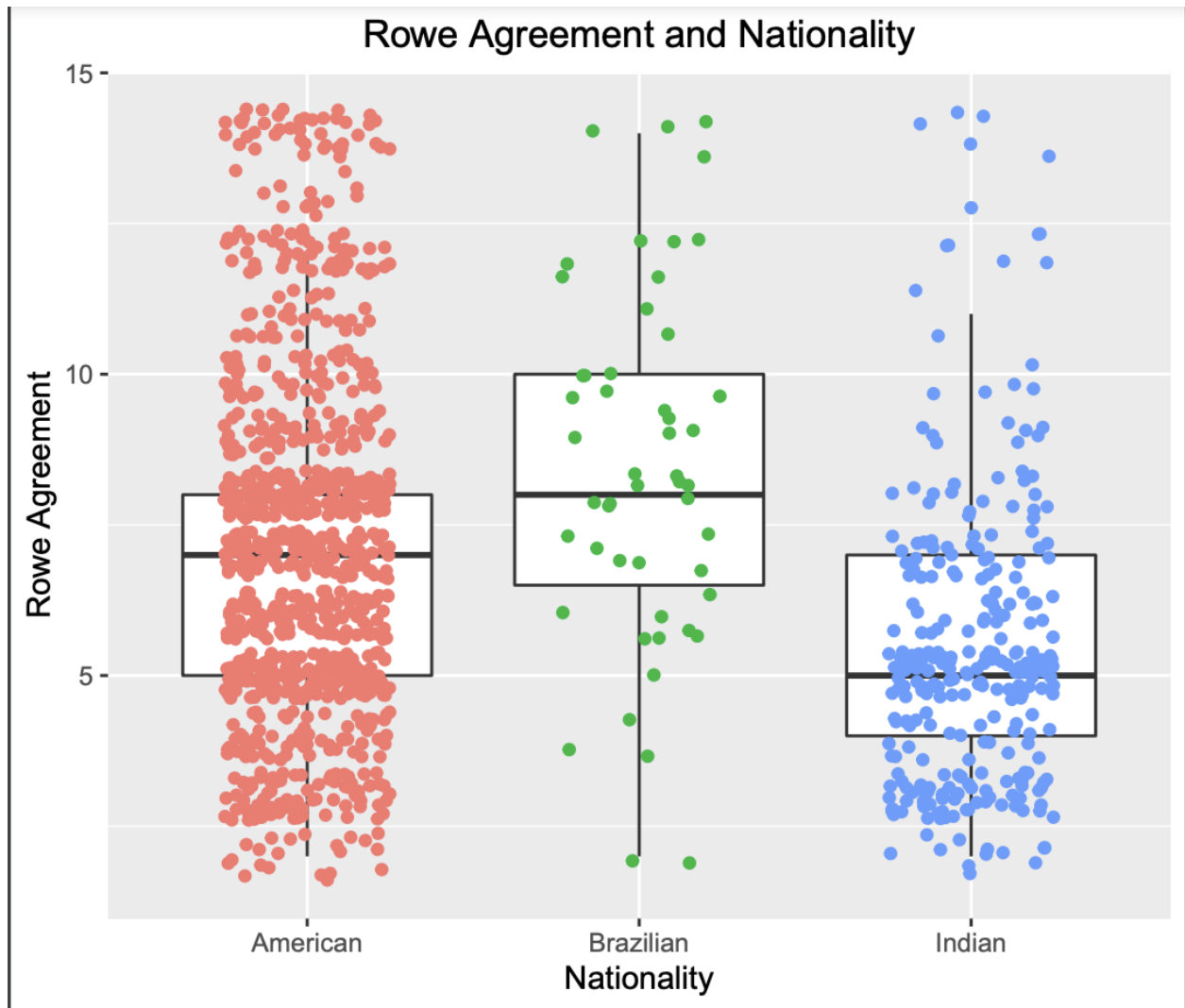
Figure 5. Rowe Agreement and Nationality

With respect to nationality, the breakdown between Agreers, Disagreers, and Agnostics was as follows:

Nationality

	Agreers	Agnostics	Disagreers
American	14.3%	13.4%	72.2%
Brazilian	23.5%	17.6%	58.8%
Indian	5.0%	5.0%	90.0%

Context, Animal-Type, and Pictures: A 2 (context) x 3 (animal-type) x 2 (pictures) ANOVA examining the relationship influence of context, animal-type, and pictures on agreement with Rowe was conducted and yielded significant results. The main effects of the type of animal or the presence of a picture were not significant, however, there was a significant main effect of context ($F_{(1)}$ = 114.303, p < 0.001, η^2 = .077, Adjusted η^2 = .070). No interaction effects were statistically significant. [Figure 6]
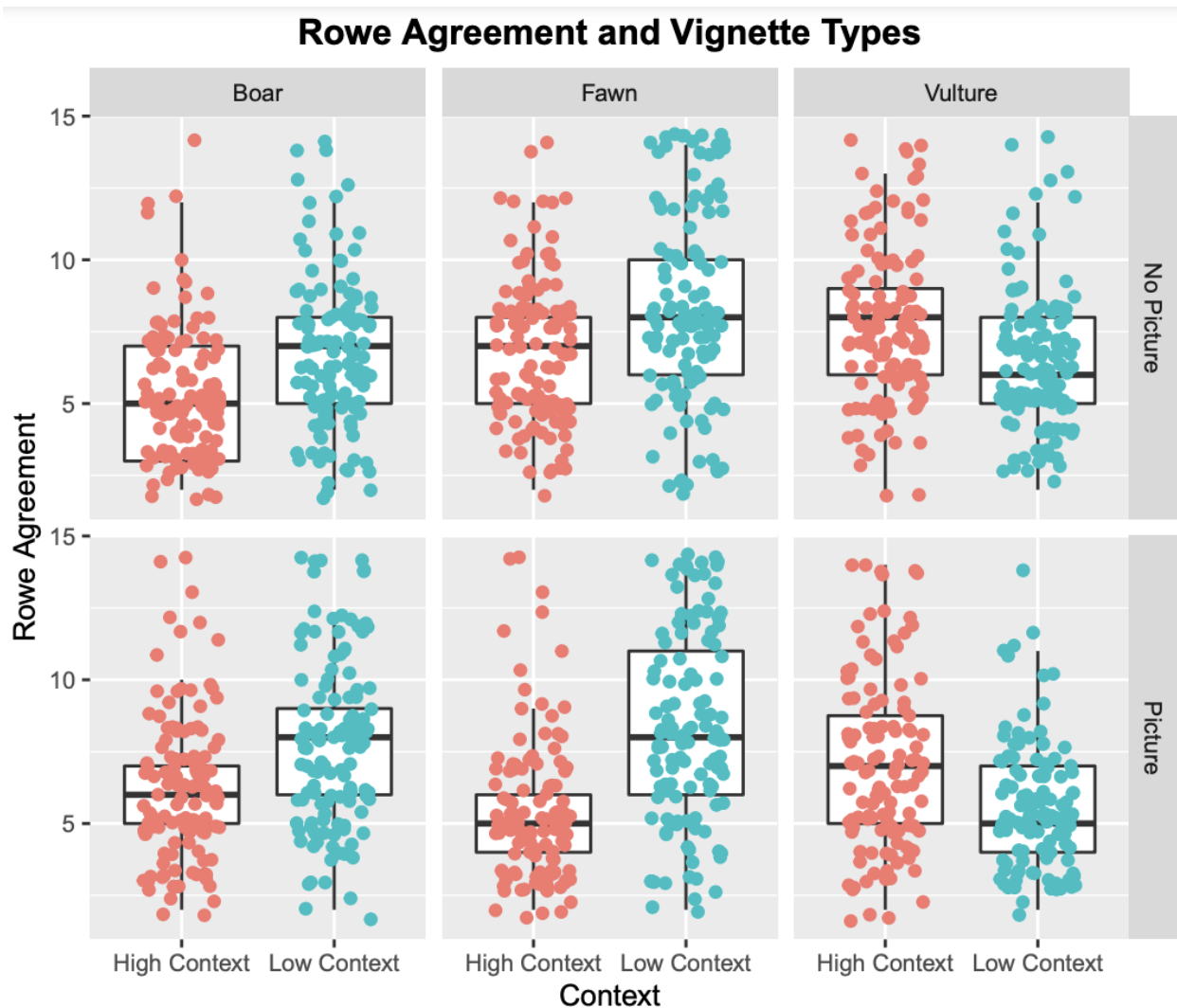

Figure 6. Rowe Agreement and Vignette Type

In sum, participants were exposed to the fawn vignette in different forms, of which the high and low context groups yielded highly significant results. The addition of context correlated with less agreement with Rowe. The more context is added, the less likely participants were to agree with

Rowe. 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 ($p = 0.022$).[10]

With respect to context, the breakdown between Agreers, Disagreers, and Agnostics was as follows:

Context

	Agreers	Agnostics	Disagreers
Low Context	21.8%	14.9%	63.4%
High Context	3.5%	9.0%	87.5%

Hunting/Butchering: To test the hunting/butchering hypothesis, Church et al. asked participants to answer the following question: "How frequently do you kill or butcher animals for food?" Participants could answer by selecting, "rarely", "sometimes", "frequently", or "very frequently." Church et al. found that overall agreement with Rowe significantly negatively correlated with the frequency of killing or slaughtering animals ($r = -0.289, p < 0.001$). In other words, the more exposure an individual had to animal death, the more likely she was to disagree with Rowe. Furthermore, when context was added to the vignette, the negative correlation between hunting/butchering and Rowe agreement was slightly stronger ($r = -0.293, p < 0.001, r_s = -0.304$ significant at the 0.01 level, 2-tailed).[11]

---

[10] For more on this result, see "The Context of Suffering" by Church, Warchol, and Barrett (2021).
[11] This Pearson correlation coefficient reflects the average of our "Low" and "High" context variables. Disagreement with Rowe decreased with the addition of context to the fawn vignette, however this initial analysis shows that the difference is not significant. This result is tentative, however, as we are running further analyses to figure out more precisely what might be going on.
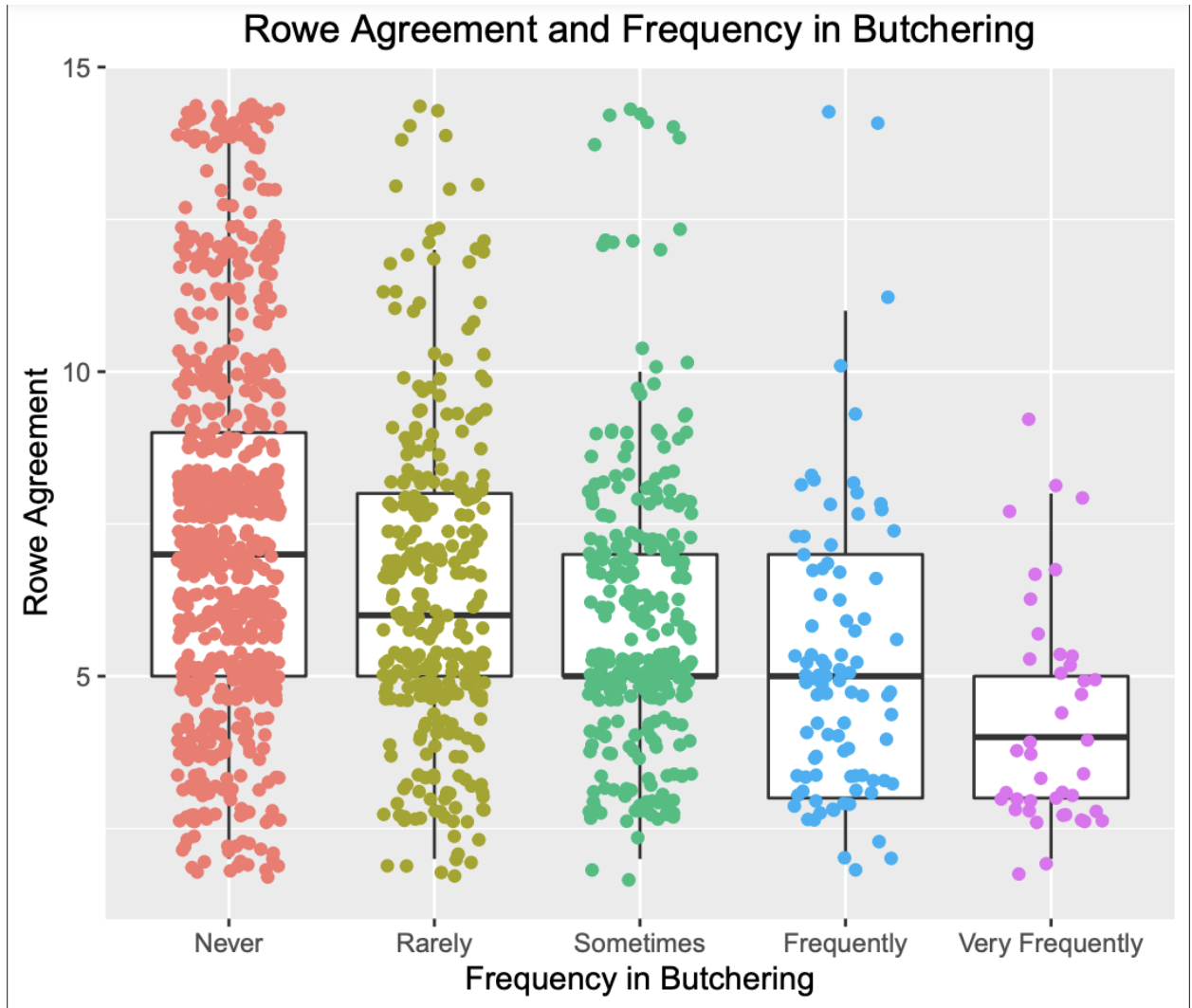
Figure 7. Rowe Agreement and Frequency in Butchering

3. Discussion

These results are surprising. It's not surprising that *some* people disagree with Rowe. After all, Rowe himself seemed perfectly cognizant of the fact that many people won't see the target cases of suffering as genuinely pointless. This was part of what motivated Rowe's "friendly" brand of atheism. That being said, Rowe reports that, among philosophers discussing the problem of evil, "there appears to be near universal agreement" that the fawn's suffering will at least *seem* pointless to most people (Rowe 2006, 79). But these findings appear to contravene these "near universal" expectations: it does *not* seem to most people that the fawn's suffering is pointless.

On second thought, however, how surprised should we really be? There were already good empirical reasons for thinking that many people wouldn't agree with Rowe. We're thinking especially of the well-established body of psychological literature showing that humans are inclined towards promiscuous teleological attributions—we can't help but see and attribute purpose, reason, and agent-causation everywhere from natural phenomena to inanimate objects. Children, for example,

manifest this tendency when explaining that rocks are "pointy" so that animals won't sit on them (Kelemen and Diyanni, 2005; Kelemen, 1999a, 1999b, 2004). In thinking that pointy objects exist for this reason, children engage in "pointful" reasoning–that is, they assume features of reality have points or reasons for their existence. They make these judgments irrespective of whether the features of reality do, in fact, exist for the reasons they cite. This indiscriminate "pointful" thinking has come to be known as *promiscuous teleology*. Evidence suggests that this tendency persists even into adulthood (Kelemen and Rosset, 2009; Kelemen, Rottman and Seston, 2013; Lombrozo, Kelemen and Zaitchick, 2007). And even scientists have been observed to make promiscuous teleological judgments under experimental duress conditions (Kelemen et al 2013). A tendency for promiscuous teleology may have been selected for its conferral of a fitness advantage early in our evolutionary history. For example, some cognitive scientists of religion argue that we evolved a hyperactive agency detection device (HADD) which inclined us to overattribute agency and avoid potentially dangerous situations (Barrett 2004). Interpreting the rustle in the bush as a predator is more conducive to fitness than assuming it's just the wind. All this to say that humans evolved to attribute agency, reasons, and general pointfulness to the world around us. This does not mean that teleological thinking is correct or even justified, but the mere descriptive point is that we are, in fact, cognitively and psychologically constituted with an orientation towards seeing purpose and agency in the natural world. And this should suggest that our findings are not as surprising as one might initially think.

In any case, what do these results mean for the overall success of Rowe's argument (again, measured by *how many* receive support for atheism via this argument and *how much* support they receive)? While the full theoretical import of these results are not entirely clear, proponents of Rowe's arguments have initial reason to worry. Consider the following results which raise questions about the argument's success.

3.1. *How Common and Confident is Agreement with Rowe?*

One of the more notable results is how few people, only 12.61%, share any level of agreement with Rowe about the gratuitousness of the fawn's suffering. Strong agreement with Rowe is rare, with only 3.9% of participants strongly agreeing that no greater good or worse evil could be achieved or prevented by the fawn's suffering. Far more people, 75.43% of all participants, reported some level of disagreement with Rowe (with the remaining 11.95% being agnostic). Indeed, it may be the case that agreement within the general population is even less prevalent than our results indicate. Atheists and Agnostics were overrepresented in our surveys compared to the general population, as is common in MTurk (see Lewis et al. 2015), and on average, these groups agreed with Rowe more than other religious identities. This is striking because the suffering of Rowe's fawn is often upheld as an especially evident instance of gratuitous evil, but our results suggest this is not the case. Indeed, if we are correct in assuming that agreement or the lack thereof is highly correlated with whether one intuits that the evil is gratuitous, then the results suggest that the fawn's suffering does not even *seem* gratuitous to most people, or at least that the intuition is not strong enough to sustain assent.

These data present a serious challenge to the overall success of Rowe's argument in at least two ways. First, they challenge the *reach* of the argument. Only those who agree with Rowe are likely to receive any rational support on the basis of this argument, and that turns out to be only a few—far fewer than one might have guessed given the argument's reputation as one of the most serious challenges to God's existence. Second, the data brings into question the *strength* of the argument, even for those who do agree with Rowe. For one, the agreement with Rowe's argument that does exist is generally moderate, and thus so is the potential level of rational support provided by it. For another, the fact that the strong majority of people disagree with Rowe should plausibly

diminish whatever level of rational confidence one might have had to begin with. In sum, these particular findings bring into question both the *extent* and the *severity* of the problem that Rowe's argument poses for theism.

In this section (3.1), we will focus more on the reach of Rowe's argument. In the next section (3.2), we will turn our attention to its strength and, in particular, whether the results provide any defeaters for those who initially agreed with Rowe.

The reach of Rowe's argument may not be as limited as the results initially suggest, for three main reasons. The first is that there may be some who *should* agree with Rowe even if they do not. That is, there may be some participants whose evidence supports the gratuitousness of the fawn's suffering but, for whatever reason, do not reach that conclusion. As an example, it may be that Theists are highly motivated to disagree with Rowe. (Indeed, as we saw in Figure 4, religious beliefs strongly correlate with how respondents evaluated the target cases.) If so, it is possible that their lack of agreement is not indicative of any failure in Rowe's argument but instead irrationality on the part of the participants. Perhaps their *evidence* agrees with Rowe, even if the participants themselves do not. If this is right, then the reach of Rowe's argument may be greater than the data initially suggests.

There will be difficulties in arguing for this though. For one, such arguments must contend directly with skeptical theism, which argues that our evidence does not permit us to draw any firm conclusions about the gratuitousness of observed sufferings. There's no bypassing such arguments by an appeal to intuition since (as we infer from our data) the participants at issue do not appear to possess such intuitions, or at least not strong ones. And of course it is open to skeptical theists and other critics of Rowe's argument to argue things the other way around—that there are some who agree with Rowe that should not. To parallel our earlier example, the motivations of Atheists and Agnostics could be brought into question just as readily as those of Theists. Finally, even if one successfully argues that some participants are being irrational by failing to agree with Rowe, one would have to show that this irrationality is extensive (significantly more so than those who irrationally agree with Rowe) in order to shift things considerably. Perhaps this burden can be shouldered, but it won't be easy to do.

The second reason to doubt whether Rowe's argument is as limited as the results initially indicate is that Church et al. only tested one example of a potentially gratuitous evil. Although Rowe's fawn is a seminal case, maybe other examples would evoke more and stronger intuitions of gratuitousness and, subsequently, agreement that the relevant evil is gratuitous. For example, Rowe's fawn is an instance of animal suffering as opposed to human suffering. Would a vignette about human suffering elicit more and stronger agreement that the evil is gratuitous? What about different kinds of human suffering? What about the suffering of a particular group of humans such as children? We don't know how many would agree with the gratuitousness of the evils presented in these various cases. And as Rowe says, "even if it should somehow be reasonable to believe [that there is a greater good made possible by or worse evil prevented by] the fawn's suffering, we must then ask whether it is reasonable to believe either of these things of *all* the instances of seemingly pointless human and animal suffering that occur daily in our world. And surely the answer to this more general question must be no" (Rowe 1979, 337).

Finally, one could contest whether the experimental design of Church et al. best captures one's level of agreement with Rowe. There are, of course, general concerns about whether participants were paying attention, whether the study will replicate, etc. More pressing is the concern

that the questions asked do not target the proper intuition.[12] Participants were asked to what extent they agreed with the following statements:

> "Some equal or greater evil could have been prevented because of the situation in the story."
> "Some equal or greater good could be accomplished because of the situation in the story."

Compare those with the following statements:

> "Some equal or greater evil could have been prevented because of the situation in the story that could not have been prevented in any other way."
> "Some equal or greater good could be accomplished because of the situation in the story that could not have been accomplished in any other way."

While these latter statements might still be too technical for a broad audience, they emphasize the necessity of the fawn's suffering–that the same goods or evils brought about or prevented by the fawn's suffering could not have been brought about or prevented through other means. And it is precisely the necessity of the fawn's suffering that Rowe is denying when he calls it "gratuitous." It may be true that the fawn's suffering is *in fact* used to accomplish some greater good or prevent some worse evil, while also being true that the fawn's suffering isn't *necessary* to accomplish that greater good or prevent that worse evil (envision a good accomplished by the fawn's suffering that could have been brought about in a different way.) Thus, it could be argued that the more accurate way to gauge what people think about the gratuitousness of the fawn's suffering is by asking whether they agree with the second set of statements rather than the first.[13]

There is something to this criticism. In avoiding talk of necessity, Church et. al. have made the questions clean and accessible, but have also sacrificed something by way of precision. It's not entirely clear that they are getting at the right intuitions to accurately measure agreement with Rowe. This is something we plan on exploring in future testing. That being said, we suspect that what Church et. al. have measured *is* a fairly reliable, though imperfect, guide to agreement with Rowe. There are sure to be some who would agree with the former statements but not the latter, but the percentage would have to be quite sizable to overturn the significant level of disagreement observed by Church et al., and that seems unlikely. Regardless, this is a matter that must be empirically probed. Until then, we will proceed under the tentative assumption that Church et al. have successfully measured the level of agreement with Rowe, at least to some significant degree of approximation.

In summary, there are several reasons to hesitate before drawing firm conclusions about the reach of Rowe's argument from this data. This is why our thesis is not that Rowe's argument is undermined, but rather that these findings issue a prima facie challenge to Rowe's argument. And the remarkable lack of agreement with Rowe about the fawn—a supposedly paradigm instance of gratuitous evil–certainly does that.

3.2. Possible Defeaters for Agreement with Rowe

---

[12] The following concerns were inspired by the comments of an anonymous referee. We are thankful to the reviewer for bringing this objection to our attention.

[13] And even then, we'd need to be careful to ensure that the operative sense of necessity is broadly logical necessity rather than necessity of a more ordinary sort. For instance: Is chemotherapy necessary to rid this person of cancer? Perhaps it is in an ordinary sense, but not in the broadly logical sense (God could miraculously cure the person without chemotherapy).

So far, we have questioned the reach of Rowe's argument. We have also suggested that, given the lack of strong agreement, the strength of that argument may not be especially formidable to begin with. In this section, we consider additional challenges to the strength of Rowe's argument. That is, we consider whether any of the reported findings pose a defeater for one's justification for agreeing with Rowe. Our initial judgment is that some of the findings, if accurate, do provide such defeaters, although this conclusion is once again subject to further testing.

Our basic line of reasoning is as follows. We see significant variation in agreement with Rowe, both overall and among various subgroups(indeed, if there is any kind of consensus, it is around those who do not judge the fawn's suffering to be gratuitous.) In most cases, those who agree with Rowe have no reason to think that their group is in a superior epistemic position compared to the others—there is, as far as they can tell, epistemic parity between the groups. And consistent with the extant experimental philosophy literature, such epistemic parity might *minimally* raise doubts about the theoretical import of the target intuitions. But in those cases where there is some reason to privilege one group over another, we will argue that the most plausible assessment tends to favor those who *don't* agree with Rowe. In either case, one's confidence that the fawn's suffering is gratuitous should plausibly be curbed.

The degree to which it is curbed depends on several factors. For one, it will depend on one's evidence concerning the relative epistemic positions of the groups at issue. Does one have any reason for privileging the judgment of one group over another? To what degree? The devil is in the details, of course, and much will hinge on how exactly we measure the strength of one's epistemic position as well as what sort of evidence justifies the subject's judgment vis-à-vis strength of epistemic position. But it seems such details are best discussed on a case by case basis than in the abstract. Thus, let us proceed to concrete instances of group variation.

3.2.1. Overall Disagreement

The first thing to consider is whether the fact that most people do not agree with Rowe should, in and of itself, give those who agree with Rowe reason to pause. Plausibly so, though it depends on whether one who agrees with Rowe has any reason to privilege one's own judgment over those who disagree. Is there some reason for thinking that *he* (or she) is better positioned to make this sort of judgment than the others who judge differently? It is certainly possible that one may have such reasons; however, we cannot think of any reasons that would be so widespread as to be available to most of those who find themselves in agreement with Rowe. Accordingly, it seems that most people who agree with Rowe will find themselves in the following conditions: they have justification for believing that more people disagree with Rowe about the gratuitousness of the fawn's suffering, and that (by their own metrics) these other people are in just as good an epistemic position with respect to whether the fawn's suffering was gratuitous or not. Arguably, this defeats their justification for thinking that the fawn's suffering is gratuitous.

What should the ultimate effect of this defeater be? At minimum, the defeater should reduce one's level of confidence that the fawn's suffering is gratuitous, moving one from one level of agreement to some weaker level of agreement. But it is also possible that the defeater should move one into a state of agnosticism about the gratuitousness of the fawn's suffering. The former is a partial defeat, the latter a full defeat. In the case of partial defeat, the strength of Rowe's argument from evil is diminished but not entirely eliminated. Because one still reasonably agrees with Rowe's premises, that argument still provides some rational support for atheism—just not as much as it did before. In the case of full defeat, however, Rowe's argument may be entirely neutralized. Agnosticism about the fawn's suffering may lead to agnosticism about the existence of gratuitous

evils (assuming that the latter is inferred from the former), preventing Rowe's argument from going through.

We take no stand on whether the predominance of disagreement with Rowe fully or only partially defeats belief that the fawn's suffering is gratuitous. It will depend on the unique evidential situation of the participant, as well as controversial issues in the epistemology of peer disagreement. But we think such findings will, for most people, result in at least some level of defeat. All of this could become moot, of course, if we discover that agreement with Rowe is (or should be) significantly more prevalent than our findings suggest. Or if new data reveals some reason to privilege the judgment of those who agree over those who disagree (perhaps we find that the cognitive processes driving the former are more reliable). So our conclusion remains tentative.

3.2.2. Nationality, Ethnicity, and Gender

Thus far, we have considered only the overall levels of agreement and disagreement. We will now focus on variation between subgroups to see if any of those results threaten or bolster the success of Rowe's argument.

Recall the following results:

1. Indians agreed with Rowe significantly less than Americans and Americans significantly less than Brazilians.
2. Asians and Black participants agree with Rowe significantly less than White and Hispanic or Latino participants.
3. Men agree with Rowe slightly (though significantly) less than women.

What should those who agree with Rowe make of these results? It seems reasonable to assume that, as a starting point, one's evidence will support epistemic parity. That is, one's evidence will indicate that the people in all of these groups are in equally good epistemic positions to discern the gratuitousness of the fawn's suffering. Objecting forces one into the exceedingly uncomfortable position of maintaining that, right off the bat, the judgment of certain nationalities, ethnicities, or genders is to be favored over that of others—a position that smacks of epistemic injustice (Fricker 2007). Of course, it should not be considered inherently prejudicial to suggest that one of these groups is epistemically superior to another in this particular kind of assessment, but you need to have the evidence to back it up. And the problem is that all of the ready explanations for privileging a group that agrees more with Rowe over one that agrees less seem unpersuasive. For instance, differences in cognitive style—specifically the tendency of those with an Eastern cognitive style to think about situations more holistically (e.g., Choi et. al. 2007)—may play some role in why Indians so strongly disagree with Rowe. This would not, however, place them in an inferior epistemic position to those with more Western cognitive styles. Just as plausibly, it places them in a *superior* epistemic position. If indeed there is parity between such groups (by the agreer's own lights mind you), then that's going to diminish whatever rational support one might have had for thinking that the fawn's suffering was gratuitous. Plausibly, it eliminates it entirely.

Something else worth noting is that within all of these groups, even Brazilians, there are far more disagreers with Rowe than agreers. Thus, even if an agreer *did* have some evidence for the epistemic superiority of, say, women over men, that would apparently only serve to lessen the degree of defeat. For example, imagine you are a woman who agrees with Rowe that the fawn's suffering is gratuitous. You learn that men on average think differently. No matter, you have evidence that women are better positioned to discern such matters. Then you discover that women *also* disagree with Rowe (73.3% disagree with Rowe; only 14.5% agree), just less so than men (76.7% disagree;

11.6% agree). Unless you have reason for privileging certain women over others, the net effect of this exchange is still to diminish one's support for thinking that the fawn's suffering is gratuitous. It just doesn't diminish it quite as much as it would have if you considered both men and women to be epistemic peers.

3.2.3. Contextual Sensitivity
Recall from above that giving context to Rowe's story significantly decreased agreement that the fawn's suffering was gratuitous. This means a fair number of those who agree with Rowe apparently do so only because certain facts about the broader context go unmentioned. We think this should make one hesitant about placing too much weight on low-context agreement.

We could flesh out this hesitancy in at least two ways. The stronger worry is that, initially, it is most reasonable to believe that those with more context are in a *better* epistemic position with respect to discerning gratuitous evil than those with less context. The more moderate worry is that, as far as we can tell, high context groups are in *at least as good* an epistemic position with respect to discerning gratuitous evil than low context groups. We'll return to the moderate worry below. For now, consider the stronger worry that, as things stand, we should privilege high context assessments over low context ones.

Privileging high context assessments makes some sense if only for the simple reason that those with more context may see potentially justifying reasons that others fail to consider. By way of analogy, if one is trying to discern the significance of a friend's action—say John lies to his wife Sarah about where he will be this Friday—one generally wants *more* context about the situation, not less. Why? Because there could be a plausible explanation that you miss without knowing the full situation. Maybe Friday is Sarah's birthday and John intends to surprise her. Even if the additional context does not itself contain any justifying reason, attending to additional context may prompt one to consider a broader range of explanations; or it may prompt one to consider *that there are* a broader range of explanations, including ones that we may not be in a position to immediately discern. It seems that one's intuitions are likely to be sharpened by such considerations if anything.

If all this is correct, then there is a defeater for low-context intuitions concerning the gratuitousness of evil. The effect of this on the overall success of Rowe's argument would be considerable. Note that most of the evils we hear about in the world—including many of the horrendous ones most liable to trigger intuitions of gratuitousness—are gathered from news headlines and other low context environments similar to the vignette about Rowe's fawn.[14]

Of course, it remains possible that low context groups are in an equal or even superior epistemic position to high context groups. Perhaps the context acts as white noise, distracting participants from honing in on the salient features of the case; or perhaps the context activates our *pareidolia*—the human tendency to discern patterns where there are none. What we really need is more data, data that helps us isolate *why* context matters. If we know more about how context is functioning in these cases, it might help us assess whether the addition of context aids or hinders one's judgment. In the meantime, however, we must decide which position—high-over-low, low-over-high, and parity—is most plausible given the information at hand. And we think that high-over-low will be (for most people) the most plausible starting point. If we are right, then those who agree with Rowe in low context conditions plausibly have a full defeater.

For the sake of argument, however, let us assume parity between low and high context groups—that is, one's evidence suggests that both groups are in equally good epistemic positions

---

[14] In fact, skeptical theists might push that *all* observed evils are comparatively low context, even those we suffer ourself or witness first hand, in that none of us knows the full story of our existence and how that evil will feature within it—the possibility of life after death being the most notable (though not the only) area of ignorance.

with respect to discerning gratuitous evil. Then that parity *still* provides a defeater for low context judgments. For if one knows that one's judgment isn't shared by those who are just as likely to be getting at the truth, this realization should plausibly lessen or even suspend that judgment.

Finally, we must once again point out that, while those in high context groups agreed with Rowe far less than those in low context groups, even those in low context groups disagreed with Rowe on average. In fact, only 21.8% of low context participants agreed with Rowe; 63.4% disagreed. So, as we saw before, *even if* an agreer determines that low context groups are to be privileged over high context ones, that person *still* should become less confident in their agreement, since most people in that same low context disagree with Rowe.

3.2.4. Education

There is at least one other finding that threatens to diminish the success of Rowe's argument: that agreement with Rowe negatively correlates with education. That is to say, at every level, the more education one receives, the less inclined people are to agree with Rowe. This is surprising (and contravened Church et al.'s initial hypothesis) since atheism is negatively correlated with religiosity (Beit-Hallahami 2006: 313). Is there something about education that diminishes intuitions of gratuitousness, or causes such intuitions to yield agreement in fewer cases? And if so, should we consider more educated individuals to be in a superior, inferior, or equal epistemic position to their less educated counterparts?

Right away, we should note that more education does not automatically (or even usually) mean that one is in a superior epistemic position with respect to some matter of judgment. Too much learning can be bad for the brain, especially if it's the wrong kind. And the effects of education can often be attributed as much to one's membership in an institution of higher learning, or a corresponding socio-economic class, as to what one learns in the classroom. It may be that less educated individuals are not only on par with but even superior to more educated individuals in making certain judgments. The point is, whether increased education puts one in a better, worse, or equal epistemic position with respect to a particular issue will depend on how education affects one's judgment. What is it doing, and does that aid or hinder discernment of gratuitousness?

One way that education might be affecting things is by providing more educated participants with a greater range of possibilities in accounting for the fawn's suffering. Perhaps they have simply been introduced to more theodicies or potential explanations of evil. Another option is that more educated participants are better at searching and applying the information they have before reaching a conclusion. That is, they are better at bringing to mind the potential theodicies of which they are aware or at seeing the application of those theodicies to the situation at hand. A third option is that education makes one more likely to question one's intuitions about the fawn. Using Kahneman's terminology, one is more likely to subject System 1 intuitions to System 2 scrutiny (Kahneman 2011). And when such reflective scrutiny does occur, it may be that educated individuals are less confident in their abilities to discern that the fawn's suffering is gratuitous. Education can be humbling—it teaches us just how little we actually know, especially when it comes to cosmic matters like the purpose of animal suffering. All of these explanations could be playing some role here, and to varying degrees, but we think the latter ones most likely. Based on our anecdotal evidence, highly educated individuals do not respond to the question of why the fawn suffers with "I know" but rather with "Who knows?" They are quick to recognize their own limitations in assessing such matters.

If these explanations are correct, a strong case could be made that more educated individuals are in a superior epistemic position with respect to the matter at hand. Recognizing this would constitute a defeater for all but the most educated of those who agree with Rowe. And even these

most educated agreers have to consider the fact that, within that most educated group of participants, only 6.5% agree with Rowe and 83.7% disagree. Unless one has reason to privilege some of these most educated individuals over others, they too will have a defeater.

But of course, we cannot be certain that these explanations *are* correct. These are the just ones that readily present themselves to us; more may be revealed through additional testing or reflection. Still, as a first pass, it seems that these results should likely lessen or perhaps even eliminate confidence in the gratuitousness of the fawn's suffering.

3.3. Other Results

Not all of our results undermine (or bolster) Rowe's argument, though they may reveal something important about how agreement with Rowe is or isn't reached. For example, consider the finding that agreement with Rowe negatively correlates with the frequency of hunting/butchering animals. At this point more research is needed before we can draw any firm conclusions; however, it's worth noting a few plausible hypotheses. Perhaps this hunting/butchering result points to the role of empathy in evaluating target instances of suffering. Presumably (though this would need to be verified) people who regularly hunt or butcher animals might be more desensitized (on average) to the suffering and death of animals than people who never butcher or hunt animals. This raises the possibility that subjects who score higher on empathy might agree with Rowe more on average than subjects who score lower on empathy. Call this the *empathy-hypothesis*. Given that women score higher on empathy than men (Hoffman, 1977), the fact that women registered more agreement with Rowe than men on average (despite men being more likely to be atheists or agnostics than women) might lend credence to such a hypothesis.

If the empathy hypothesis were confirmed, what conclusions could we draw? It's difficult to say. On the one hand, empathy-driven judgments can be notoriously unreliable. Empathy-driven judgements are highly variable according to factors that often seem morally irrelevant, so there's reason to think that such a faculty might not be fully veritic. That said, a wholesale skepticism regarding empathy-driven judgments would be a radical conclusion—perhaps leading to an intractable form of moral skepticism (see Haidt 2012, chapters 1 and 2). More work needs to be done regarding the kind of empathy at work in assessing the target cases of suffering before any firm conclusions could be drawn.

Whatever we want to say about the role of empathy, cuteness did *not* make as significant of a difference as Church et al. anticipated. Recent work on cuteness suggests that humans "hyper-mentalize" cute entities, and are thus more likely to anthropomorphize or attribute complex mental states and moral interests to them (Sherman & Haidt 2011). If it is the case that people hyper-menatalize cuter animals, and if empathy plays a significant role in determining agreement with Rowe, then why don't we see cuteness of the animal (a fawn vs. a boar or a vulture) making more of a difference? The upshot, especially in conjunction with the empathy hypothesis, is not entirely clear.

4. Conclusion

To recap, Church, Warchol, and Barrett found that the vast majority of respondents did not agree with Rowe regarding target examples of suffering. They found that agreement varied significantly across various demographics (including gender, nationality, education level, etc.). They also found that ascriptions of gratuitousness significantly diminished when the target example of suffering was accompanied with context. We have argued that, if these results are indeed accurate, then they put those who agree with Rowe in an epistemically perilous position. Few, it seems, will have strong reasons for privileging the judgment of those who agree with Rowe over those who

don't. Accordingly, one's justification for believing that the fawn's suffering is gratuitous would be partially or even fully defeated. This would remove some justification for believing the second premise of the evidential problem of evil, possibly undermining Rowe's argument. Given the exploratory nature of the research thus far, such conclusions are highly defeasible. In particular, we discussed a reason for doubting whether Church et al. have accurately measured agreement with Rowe in section 3.1. In light of this concern, we must emphasize the conditional nature of our conclusion: *if* these results are an accurate measure of agreement with Rowe, then Rowe's problem of evil is in trouble. What we can conclude unconditionally is that these findings raise serious questions about the strength of Rowe's argument–questions that deserve further investigation.[15]

Bibliography

Alston, William P. *Perceiving God: The Epistemology of Religious Experience* Ithaca, NY: Cornell University Press, 1991a.

Alston, William P. "The Inductive Argument From Evil and the Human Cognitive Condition." *Philosophical Perspectives* 5 (1991b): 29–67. https://doi.org/10.2307/2214090.

Barrett, Justin L. *Why would anyone believe in God?*. AltaMira Press, 2004.

Bergmann, Michael. "Skeptical Theism and Rowe's New Evidential Argument From Evil." *Noûs* 35, no. 2 (2001): 278–96. https://doi.org/10.1111/0029-4624.00297.

Bergmann, Michael. "Commonsense Skeptical Theism." In *Reason, Metaphysics, and Mind*, edited by Kelly James Clark and Michael Rea. Oxford University Press, 2012. https://doi.org/10.1093/acprof:oso/9780199766864.003.0002.

Beit-Hallahmi, Benjamin. 2006. "Atheists: A Psychological Profile." In *The Cambridge Companion to Atheism*, 300–318.

Chisholm, Roderick M. *Theory of Knowledge* 2nd ed. Englewood Cliffs, N.J.: Prentice-Hall, 1977.

Choi, Incheol, Minkyung Koo, and Jong An Choi. "Individual Differences in Analytic Versus Holistic Thinking." *Personality and Social Psychology Bulletin* 33, no. 5 (May 1, 2007): 691–705. https://doi.org/10.1177/0146167206298568.

Church, Ian M., Rebecca Carlson, and Justin Barrett. 2021. "Evil Intuitions? The Problem of Evil, Experimental Philosophy, and the Need for Psychological Research." *Journal of Psychology and Theology* 49(2), 126–141.

Church, Ian M, Isaac Warchol, and Justin L. Barrett. 2021. "The Context of Suffering: Empirical Insights into the Problem of Evil." *TheoLogica: An International Journal for Philosophy of Religion and Philosophical Theology*.

Dougherty, Trent. "Epistemological Considerations Concerning Skeptical Theism." *Faith and Philosophy* 25, no. 2 (2008): 172–76. https://doi.org/10.5840/faithphil200825215.

Dougherty, Trent. "Further Epistemological Considerations Concerning Skeptical Theism." *Faith and Philosophy* 28, no. 3 (2011): 332–40. https://doi.org/10.5840/faithphil201128331.

Draper, Paul. "Evil and the Proper Basicality of Belief in God." *Faith and Philosophy* 8, no. 2 (1991): 135–47. https://doi.org/10.5840/faithphil19918215.

Fricker, Miranda. 2007. *Epistemic Injustice: Power and the Ethics of Knowing* 1 edition. Oxford ; New York: OUP Oxford.

Gellman, Jerome. "The Experience of Evil and Support for Atheism." In *The Blackwell Companion to the Problem of Evil*, edited by Justin McBrayer and Daniel Howard-Snyder, 1st edition. John Wiley & Sons, 2014.

Gellman, Jerome. "A Surviving Version of the Common Sense Problem of Evil." *Faith and Philosophy* 34, no. 1 (2017): 82–92. https://doi.org/10.5840/faithphil2016122971.

Haidt, Jonathan. 2012. *The Righteous Mind: Why Good People Are Divided by Politics and Religion*. New York: Vintage.

Hasker, William. "The Necessity of Gratuitous Evil." *Faith and Philosophy* 9, no. 1 (1992): 23–44. https://doi.org/10.5840/faithphil1992911.

Hendricks, Perry. "The *Real* Commonsense Problem of Evil." n.d.

Huemer, Michael. *Skepticism and the Veil of Perception* Lanham, MD: Rowman & Littlefield, 2001.

Huemer, Michael. "Compassionate Phenomenal Conservatism." *Philosophy and Phenomenological Research* 74, no. 1 (2007): 30–55. https://doi.org/10.1111/j.1933-1592.2007.00002.x.

Hoffman, Martin L. 1977. "Sex Differences in Empathy and Related Behaviors." *Psychological Bulletin* 84 (4): 712–22. https://doi.org/10.1037/0033-2909.84.4.712.

Inwagen, Peter van. "The Magnitude, Duration, and Distribution of Evil: A Theodicy." *Philosophical Topics* 16, no. 2 (1988): 161–87. https://doi.org/10.5840/philtopics198816218.

Inwagen, Peter van. "The Problem of Evil, the Problem of Air, and the Problem of Silence." *Philosophical Perspectives* 5 (1991): 135–65. https://doi.org/10.2307/2214093.

Inwagen, Peter van. "The Argument From Particular Horrendous Evils." *Proceedings of the American Catholic Philosophical Association* 74 (2000): 65–80. https://doi.org/acpaproc20007429.

Kahneman, Daniel. 2011. *Thinking, fast and slow*. Macmillan.

Kelemen, Deborah. (1999a). The Scope of Teleological Thinking in Preschool Children. Cognition, 70, 241-272.

Kelemen, Deborah. (1999b). Why are Rocks Pointy? Children's Preference for Teleological Explanations of the Natural World. Developmental Psychology, 35, 1440-1452.

Kelemen, Deborah. (2004). Are Children "Intuitive Theists"? Reasoning about Purpose and Design in Nature. Psychological Science, 15, 295-301.

Kelemen, Deborah. (2012). Teleological Minds: How Natural Intuitions About Agency and Purpose Influence Learning About Evolution. In K. S. Rosengren, S. K. Brem, E. M. Evans & G. M. Sinatra (Eds.), Evolution challenges: Integrating research and practice in teaching and learning about evolution. Oxford: Oxford University Press.

Kelemen, Deborah. & DiYanni, C. (2005). Intuitions about Origins: Purpose and Intelligent Design in Children's Reasoning about Nature. Journal of Cognition and Development, 6, 3-31.

Kelemen, Deborah., & Rosset, E. (2009). The Human Function Compunction: Teleological Explanation in Adults. Cognition, 111, 138-143.

Kelemen, Deborah., Rottman, J. & Seston, R. (2013). Professional Physical Scientists Display Tenacious Teleological Tendencies: Purpose-Based Reasoning as a Cognitive Default. Journal of Experimental Psychology: General, 142, 1074-1083.

Kelly, Daniel., 2011. *Yuck!: The nature and moral significance of disgust* MIT press.

Lombrozo, Tania., Kelemen, Deborah., & Zaitchik, D. (2007). Inferring Design: Evidence for a Preference for Teleological Explanation in Patients with Alzheimer's Disease. Psychological Science, 18, 999-1006.

Peterson, Michael L. *Evil and the Christian God*. Grand Rapids, Mich: Baker Book House, 1982.

Matheson, Jonathan D. "Epistemological Considerations Concerning Skeptical Theism: A Response To Dougherty." *Faith and Philosophy* 28, no. 3 (2011): 323–31. https://doi.org/faithphil201128330.

Moore, George Edward. "Proof of an External World."*Proceedings of the British Academy*25, no. 5 (1939): 273–300.

Moore, G. E. "A Defence of Common Sense." In*Philosophical Papers*. New York, NY: Macmillan, 1959.

Plantinga, Alvin. 2000. *Warranted Christian Belief*. New York: Oxford University Press.

Plantinga, Alvin, and Nicholas Wolterstorff. *Faith and Rationality: Reason and Belief in God* University of Notre Dame Press, 1983.

Pryor, James. "The Skeptic and the Dogmatist." *Noûs* 34, no. 4 (2000): 517–49. https://doi.org/10.1111/0029-4624.00277.

Rowe, William L. 1979. "The Problem of Evil and Some Varieties of Atheism."*American Philosophical Quarterly* 16 (4): 335–41.

Rowe, William L. "Ruminations About Evil." *Philosophical Perspectives* 5 (1991): 69–88. https://doi.org/10.2307/2214091.

Rowe, William L. "Friendly Atheism, Skeptical Theism, and the Problem of Evil." *International Journal for Philosophy of Religion*59, no. 2 (2006): 79–92. https://doi.org/10.1007/s11153-005-6178-6.

Rutledge, Jonathan Curtis. "Perspectival Skeptical Theism." *Faith and Philosophy* 36, no. 2 (2019): 244–64. https://doi.org/10.5840/faithphil2019430121.

Senor, Thomas. "Skeptical Theism, CORNEA, and Common Sense Epistemology." In *The Blackwell Companion to the Problem of Evil,* edited by Justin McBrayer and Daniel Howard-Snyder, 1st edition. John Wiley & Sons, 2014.

Shields, Christopher. "The Phainomenological Method in Aristotle's Metaphysics." In *Aristotle on Method and Metaphysics*, edited by Edward Feser, 7–27. Philosophers in Depth. London, UK: Palgrave Macmillan, 2013.

Sherman, Gary D., and Jonathan Haidt. 2011. "Cuteness and Disgust: The Humanizing and Dehumanizing Effects of Emotion."*Emotion Review* 3 (3): 245–51. https://doi.org/10.1177/1754073911402396.

Stump, Eleonore. *Wandering in Darkness: Narrative and the Problem of Suffering* Oxford University Press, 2010.

Swinburne, Richard. *Epistemic Justification*. Oxford, UK: Oxford University Press, 2001.

Swinburne, Richard. *The Existence of God* Oxford University Press, 2004.

Tweedt, Chris. "Defusing the Common Sense Problem of Evil."*Faith and Philosophy* 32, no. 4 (2015): 391–403. https://doi.org/10.5840/faithphil201592347.

Wykstra, Stephen J. "The Humean Obstacle to Evidential Arguments From Suffering: On Avoiding the Evils of 'Appearance.'"*International Journal for Philosophy of Religion*16, no. 2 (1984): 73–93. https://doi.org/10.1007/bf00136567.

Wykstra, Stephen J. "Rowe's Noseeum Arguments From Evil." In *The Evidential Argument From Evil*, edited by Daniel Howard-Snyder, 126–50. Indiana University Press, 1996.