Dear Dr. von Hippel,

Thank you for your attention to our manuscript. Your feedback, along with feedback from both reviewers, was very helpful. We have now revised the manuscript quite heavily to clarify a number of points. Our main changes and clarifications are now highlighted in yellow in the revision (many of our changes were deletions and are described here). Broadly, we made a few key changes:

1) **We have substantially reworked the Introduction to lead with the different theories of religious belief and disbelief that we are comparing. As you mention, the original Intro was quite brief and the Discussion quite long. We have flipped these priorities in an attempt to really get the framing clear at the outset. Pages 4-6 now explicitly describe in more depth each of the three primary theoretical approaches we juxtapose.**
    a. **"Different academic subfields and traditions have converged on a few primary theoretical approaches for understanding religious belief and disbelief. Three of the most prominent are secularization theories, the cognitive byproduct approach made popular by evolutionary psychology and the cognitive science of religion, and a dual inheritance approach derived from work in cultural evolution and gene-culture coevolution." This section precedes subsections with more discussion of each approach. After this, predictors of atheism are introduced, and then finally the Intro outlines hypotheses each theoretical approach makes about predictors of disbelief.**
2) **Fleshing out the Intro necessitated trimming some of the Discussion section. We largely removed the Metascientific Implications (different sorts of replications, etc…) section that Reviewer 1 found problematic.**
3) **We have reframed Results, tables, and figures to refer only to the variables being tested in the models, rather than the conceptual phrasing we initially used, per Reviewer 1's suggestions.**
4) **We acknowledge and discuss some of the analytic issues raised by Reviewer 1 concerning issues of measure directness and reliability and how they may impact our parameter estimates.**
    a. **P. 19: "Of the four primary atheism predictors that we used to test prominent theories, religious CREDs emerged as a clear empirical winner. In some ways, however, our tests may have been methodologically stacked in this variable's favor. Like the self-reports of religious disbelief, this measure includes self-report items about religious upbringing. Thus there is shared method variance associated with this predictor that is less evident for others. Further, predictors varied in both reliability and demonstrated validity. We chose these measures simply because they have been used in previous research; that said, previous use does not necessarily imply that the measures were sufficient. As with much of psychology, measurement quality is a concern."**
5) **Reviewer 1 noted that social liberalism was strongly associated with measures of religious disbelief. This is an interesting pattern, although it is theoretically quite**

remote from the present project. As such, we discuss it as an additional pattern that may warrant additional investigation.

a. **P. 20: "Although it is not featured in any of the core theoretical perspectives we evaluated, social liberalism was consistently the strongest covariate of religious disbelief. The intersection of religious and political ideology is an interesting topic in its own right, and merits further consideration. Interestingly, disbelief if anything was associated with fiscal conservatism in this sample. This suggests that simple 'faithfuls are conservative' tropes are oversimplifications. Ideology and religiosity are multifaceted and dissociable, but certainly of interest given rampant political polarization in the USA and elsewhere. At the same time, we caution readers that religion-ideology associations, whatever they may be, are largely orthogonal to existing cultural and evolutionary theories of religious belief and disbelief."**

Our revision is now much clearer regarding predictions made and the theoretical lessons that can (and cannot) be drawn from it. We thank you and the reviewers for clear feedback that helped us strengthen our work and make it more intelligible to researchers outside of our own immediate research bubble.

We outline our more specific responses below.

Thank you for your attention to our manuscript!
Sincerely,
[redacted author list]

Editorial Decision Letter.

26-May-2020

Dear Dr. Gervais:

Manuscript ID SPPS-20-0136 entitled "The Origins of Religious Disbelief: A Dual Inheritance Approach" which you submitted to Social Psychological and Personality Science, has been reviewed. The comments of the reviewer(s) are included at the bottom of this letter.

Let me add a little context to the reviews. First and foremost, I think the most fundamental challenge is to respond to the issues raised by Reviewer 1 in the relatively brief space you're allotted in SPPS. This is not to say that I disagree with any of Reviewer 2's points, as I don't, but that (s)he has requested less of you and hence it will be more manageable to address these concerns.

As you consider how to respond to these reviews, let me add that there were points at which I struggled to follow exactly what you were saying - not because the writing is unclear, but

because I'm not expert in the area and you have (by necessity) laid out the problem rather quickly. From my perspective, the intro seemed overly brief and the discussion overly long. If you could provide a clearer description of the problem up front, and why your approach is informative and important, then I think less will be demanded of you in the discussion section.

As detailed above, we have now substantially reworked the framing. There is a new section in the Introduction that provides a concise introduction to the theoretical approaches we are comparing (secularism, cognitive byproduct, and dual inheritance). We hope this gives readers outside of our immediate area an easier "in" to this work. To make space for this, we trimmed a lot of the less popular sections from the General Discussion.

I was also concerned about the large rate of participant exclusion in the study, as it is a potential threat to the study's validity. Perhaps you could address this concern by footnoting analyses that include everyone, whether they passed the attention check or not.

We now clarify that the exclusion rate was from a combination of participants failing a subtle attention check item, and ALSO participants who did not complete all measures used in the overall models. Our exclusion rate is, if anything, markedly lower than we typically observe in MTurk or student samples. None of our inferences hinge on the exclusions (now noted in manuscript), and we share our data freely for those who wish to explore further.

Finally, let me add that if you choose to revise and resubmit the manuscript, I might run it by Reviewer 1 again. Additionally, I will probably run it by an expert in Bayesian analyses, as that was remiss of me not to address that issue in the first submission.

Our analyses aren't anything fancy. Their guts are just basic regressions. The only shift that comes with Bayesian estimation is the interpretation of parameter estimates. As we note in our "Analytic Approach" section, the Bayesian parameter estimate outputs (posterior probabilities and HPDIs) essentially just create the inferences that people often mistakenly think they're getting from frequentist analyses. But the models are very straightforward regressions (either linear or logistic).

To revise your manuscript, log into https://mc.manuscriptcentral.com/spps and enter your Author Center, where you will find your manuscript title listed under "Manuscripts with Decisions." Under "Actions," click on "Create a Revision." Your manuscript number has been appended to denote a revision.

When submitting your revised manuscript, you will be able to respond to the comments made by the reviewer(s) in the space provided. Please also use this space to document any changes you made to the original manuscript. In order to expedite the processing of the revised manuscript, please be as specific as possible in your response to the reviewer(s).

You will be unable to make your revisions on the originally submitted version of the manuscript. Instead, revise your manuscript using a word processing program and save it on your computer. Once the revised manuscript is prepared, you can upload it in step 5 of the

revision submission process.

IMPORTANT: Your original files are available to you when you upload your revised manuscript. Please delete any redundant files before completing the submission.

Because we are trying to facilitate timely publication of manuscripts submitted to Social Psychological and Personality Science, your revised manuscript should be uploaded as soon as possible. If it is not possible for you to submit your revision in a reasonable amount of time, we may have to consider your paper as a new submission.

Once again, thank you for submitting your manuscript to Social Psychological and Personality Science. I look forward to receiving your revision.

Sincerely,
[redacted]
Guest Action Editor

Thank you for your generally positive appraisal and feedback about aspects of our paper that required clarification.

Reviewer(s)' Comments to Author:

Reviewer: 1

Comments to the Author
(Apologies for the formatting here. I had hoped to upload a document, but I need to paste into a textbox).

The present article quantifies the relationship between religiosity (operationalised as low scores using Jong et al 2013) and several proxy-measures associated with theories of belief/disbelief. The authors argue that they test several competing models that purport to explain disbelief, and find strong support for the [absence of] CREDs, corresponding with the dual-inheritance model of disbelief. Analytic disbelief (operationalized using a measure of cognitive reflection) also predicted disbelief, but to a lesser extent. Reported (male) gender identity predicted disbelief. The strongest predictor in the model was being socially liberal. The authors sampled a nationally representative group of people living in the US.

The authors employ bayesian statistics. I will admit here and now that I am not sufficiently expert to review the specifics of their analysis, and I would caution against publication until a competent reviewer can evaluate. (Though I will make some comments regarding analysis).

Overall, this is a simple and elegant paper, and represents a meaningful contribution that extends beyond the disbelief literature.

Thank you for this generally positive appraisal. We will echo that this study is probably the most comprehensive theoretically-testing analysis of the predictors of religious disbelief that has been conducted in the USA. We hope its appeal extends beyond just folks who study the cognitive and evolutionary origins of religion to people interested in evolution and culture more broadly.

As it stands, the paper can be read and largely understood as is (with the caveat that not everyone understands Bayes). However, the brevity in explaining and justifying some decisions poses a challenge, though I concede this is likely a product of the word count associated with the journal rather than a lack of thought on the part of the authors. Though I believe some of the claims made in the discussion over-state the work's significance, these do not detract from the findings (and in this spirit, I would much prefer words to be spent on explaining certain decisions rather than contextualising the results as a contribution to formal theories and non-WEIRD psychology - two things this paper is not).

We have toned down some of the meta-scientific speculation that cluttered the initial submission's Discussion section.

Major points:

All the measures used to operationalize various models, with the exception of CREDs, are indirect. Though they are treated as though are direct. This occurs in figure 1, where model-names are substituted for measurement-values. This occurs throughout the discussion. This is to confuse the map for the terrain, even though they must correspond. Please be consistent and refer to measures as measures when technically appropriate.

Thanks for this point. In our effort to link the findings to the theories being tested, we overshot and our terminology was clearly misleading. We now discuss all measures strictly as measures, and leave the conceptual labels (analytic atheism, etc.) for the Intro and Discussion sections, where they belong.

Relatedly, the CREDs items ask about religion, but the other three scales measure something trait-like about the individual generally. That cognitive reflection predicts atheism less than CREDs is not surprising, given that reflection was measuring nothing so direct as CREDs. I don't think at present, on the strength of the evidence, that one can claim one model of disbelief is better than another model of disbelief, because the measures themselves are not equivalent. This critique may, however, be unavoidable, as the models work at different levels of specificity to the individual. This will need to be addressed in some depth.

The SBS measure, and the CREDs measure had high reliability. The other scales all had low reliability (usually < .8). While heuristically satisfactory, the low reliabilities likely suggest that the true degree of predictive power is an under-estimate. This blog entry illustrates the issues with reliability, and implicitly demonstrates how the measures in the present paper - having quite varied differences in reliability - are likely to predict different amounts of variance in the outcome simply by virtue of increased measurement error.

https://medium.com/@Sam_D_Parsons/ignoring-measurement-reliability-is-a-real-life-horror-story-b98a2517db26.

I'm not sure exactly how the authors can address this. The data has already been collected, and has obvious merit. However the conclusions about the relative value of each model of disbelief do not seem as firm to me as they appear presented in the paper.

Thank you for raising these important points. We now include explicit discussion of the issues R1 raised regarding measurement reliabilities and varying directness of measures on p. 20. We note that the deck may have been somewhat stacked in favor of the CREDs measure. However, we note that even these important methodological arguments are not enough to salvage hard interpretations of, say, secularization or cognitive byproduct accounts. Byproduct accounts, in particular, have explicitly rejected the CREDs prediction and also are quite clear that cognitive reflection ought to be the primary predictor: "A common refrain in the cognitive science of religion is that atheism must necessarily require effortful cognitive reflection. Prominent scholars of this tradition claim, for example, that atheism "require[s]. . . cognitive effort" (Barrett, 2010) and that "disbelief is generally the result of deliberate, effortful work" (Boyer, 2008). Here, atheism is predicted to be rare, potentially psychologically superficial, and emerging either through deficient social cognition or superior analytic processing." Our data, combined with recent cross-cultural work and meta-analyses, do not support this prediction.

This fact, coupled with how (in)direct the measures are of the purported construct will need some dedicated focus.

To what extent are the measures correlated and co-linear? Probably not a great deal, but since the study really only has 5 focal variables, I would argue that the authors should include a correlation matrix.

Multicollinearity was very much not an issue. We now note in the manuscript that core predictor correlations were quite modest (ranging from r = -.12 to .22).

Is there a reason none of the regressions use a person's own attendance/participation in religious events as a predictor of belief? Determining a causal arrow is impossible in the present context, but it's not clear why the data was collected if it wasn't really going to be used (was it only for the readers interest in the demographics? Surely that can't be the case, because you specifically recruited a nationally representative sample.). I can't think of a sensible justification for why the authors did not use religious-participation to predict religious-(dis)belief. Certainly their pre-reg document doesn't include it, but what justification is there for ignoring such a relevant measure at that point in the planning? Certainly even atheists attend religious events at some non-0 rate (and all your variables are continuous, anyway. Except for the binary belief in god, but even so). If the authors so-choose to not run this analysis, I think they need to defend that decision. We must keep in mind that CREDs not only predicts how viewers ought to respond to perceiving others performing costly actions, but speaks to one's own motivation, ability and willingness to participate in such costly action. Relative frequency of participation itself (once a year vs. one a week) should be accounted for a feature that maintains (dis)belief.

We did not include attendance as a covariate in our analyses for a couple of reasons. First, it correlates quite strongly with belief and disbelief. For the continuous measure, for example, the correlation is ~ .6. The attendance (and prayer, for that matter) variables could sensibly be included as additional items in the belief and disbelief measure, on strictly empirical terms. Second, this correlation makes inferences problematic if it were included as a predictor. The various theories are fairly clear in their predictions about disbelief. They do not, however, make any clear predictions about what ought to be associated with the residual variance in disbelief that remains after attendance is accounted for. An analysis predicting disbelief while also accounting for the attendance differences risks making inferences about a quantity that is wholly independent of theory. At best, the contribution of such an analysis is theoretically murky. At worst, it is statistically inappropriate and theoretically misleading. We discuss this point in the manuscript, p. 11:

"We did not include these religious demographic variables in our primary statistical models because the risks of multicollinearity and redundant variance with the outcome measure far outweigh any value of such models for theory testing, the primary goal of this project. Put simply: it is not entirely clear what any prominent theory predicts about what cognitive or cultural factors are associated with the residual variance in religious belief, independent of prayer and attendance. That said, our data are open and freely shared; we encourage curious researchers to explore trends with attendance, prayer, or other religious demographics that interest them."

Put differently: if our goal was to just explain as much variance in disbelief as possible, we would absolutely include prayer, church attendance, etc. But because our goal is instead testing theoretically-consistent hypotheses, chucking in all relevant predictors to boost R^2 is more likely to mislead than inform.

The SBS regression, and the logistic regression are not independent analysis. Nor is the 'individual zero-sum replication analysis'. These cannot sensible be regarded as converging evidence. (Unless somehow bayes accounts for this?). What is the Point-Biserial Correlation between SBS scores and the binary scores? Does it justify running both analysis? And I don't understand the logic of the 'replication analysis'. You've demonstrated in a more-complete model that some variables are predictive; what value is there is running simpler models that account for less variance? (See Middling point 1)

Here and in another comment, R1 hints that we might be 'double dipping' by running parallel analyses. First, we would like to point out that the common intuition that running multiple rounds of analyses on a dataset is problematic is itself problematic and likely mistaken (see, for example, this preprint by Devezer et al: https://www.biorxiv.org/content/10.1101/2020.04.26.048306v1 which explicitly debunks the argument that data oughtn't be probed more than once). Second, we performed both the continuous and binary analyses to establish some robustness. This is pretty standard in cases (like this) where there are different ways to measure a key construct. Inferences are practically identical across two distinct operationalizations of disbelief.

The individual replication analyses were a slightly different sort of thing. Each of those predictors has been successfully used in previous research, but not simultaneously modeled. We were merely checking whether those individual associations emerged in our more representative dataset. This is an aside from our core theoretical tests (in which models have all predictors simultaneously). But it's a nice check on previous work. The full models can't sensibly answer this same question because adding predictors to a model can often drastically alter the results. Out of fairness to previous work using these variables individually, we followed their precedent in this ancillary analysis. For space reasons, we have moved these analyses to the Online Supplement.

I won't impugn the authors for using Bayes. Though it must be admitted that far fewer people understand Bayes than they do frequentist statistics, and not only is it unclear why bayes was chosen here, it's not sufficiently explained. That's not to say it's not justifiable, nor understandable. Just that it isn't justified, and not easily understood by a naive reader. Certainly, scholars who use bayesian analysis must get sick of explaining bayes, but the burden of clear communication and understanding is not solely on the reader, but on the person with the message. At a minimum the authors should link to a primer for the reader. The authors also introduce terms without explanation (such as 'golems').

We included a brief primer on Bayesian analysis and its interpretations. In this primer (pp. 13-14), we reference a number of quite accessible short papers that can walk through some toy examples and more fully illustrate the pragmatic benefits of a Bayesian approach.

Being socially liberal is the strongest predictor in the model. If I'm reading HPDI correctly (presuming analogy to CI's), the effect of being socially liberal doesn't even overlap with the effect of CREDs. This cannot simply be ignored. It might suggest that adoption of 'socially liberal values' is incompatible with religious belief. To the extent that any causation is inferred in the present analyses, this inference (or others of the author's choosing) needs to be discussed. Particularly because CREDs correspond with social learning, and the other features are cognitive. And yet the single biggest predictor in the model is a social factor (presumably capturing a constellation of beliefs and values). What if being social liberal (and everything that broadly means) just doesn't gel with US-style religion? (Certainly, being religious might ipso facto make someone identify as liberal, but even if this is the case, it needs to be discussed. Also lending support for correlation tables - how dissociable are the other predictors from being liberal? It might be the case that CREDs are orthogonal, but cognitive reflection is not).

R1 correctly notes that religious disbelief is associated with social liberalism. This is an interesting pattern and may have emerged for various US historical, social, or even cognitive reasons. Alas, these reasons are quite theoretically remote from the goals of the present paper, as none of the prominent theories we compare make any predictions about the relationships between disbelief and political ideology in the USA. Perhaps this is because they take a broader evolutionary approach. That said, we discuss the political findings (p. 20), see section quoted above.

Middling points:

The authors are pretty loose with the term 'replication'. In part III they double-dip on analysis. Had the individual analyses been presented first (demonstrating validity of your measures and generalizability from the literature), then combined into your focal analyses… then maybe this would be persuasive. But right now you present the focal analysis, make inferences about which models of disbelief are valid, then fail to demonstrate previously published effects. Either you replicated the individual effects and found them valid, thus allowing you to reject them as plausible explanation for disbelief, OR, you failed to replicate the effects and deem them invalid primae facie, but if so, you can't then justify their contribution to your focal analysis. It strikes me that a) they need to be in your focal analysis, because precedent demands it, and b) failure to find an effect doesn't mean the variables aren't contributing to the overall model in important ways. In part III, the authors claim: That one of the candidate factors culled from existing literature did not appear as a robust predictor may suggest tempered enthusiasm for its utility as a predictor of individual differences in religiosity more broadly. But this claim is not more valid for being tested in isolation than it was when tested in a more complete model (assuming no major issues of colinearity).

In our opinion, R1 is being too harsh when dismissing multiple converging analyses as 'double dipping.' First, we note that the common intuition about multiple analyses on a dataset being problematic is in all likelihood not true (https://www.biorxiv.org/content/10.1101/2020.04.26.048306v1). Second, we are most definitely not 'double dipping' in the most problematic sense – namely deriving and subsequently testing hypotheses on the same set. We just present 2 converging analyses evaluating robustness across slightly different operationalizations of the key outcome measure, followed by straightforward analyses trying to replicate published patterns as an aside. There is nothing statistically fishy about this. And, it is quite common for associations between variables to change when testing predictors in isolation and testing them simultaneously in multiple regression. That things were pretty similar using slightly different approaches, in models with and without relevant covariates, in our work shows convergence and robustness.

That said, we recognize that our discussion of replication in the GD set off some alarm bells. So we removed most discussion of replication from the Discussion, and relocated the individual variable analyses to the Online Supplement.

In the discussion the authors conflate 'cognitive reflection' with 'analytic atheism'. Surely one could only make the claim that it's analytic atheism if one measured the actual reasons for why one holds their belief (God is [not] real, and I am convinced by the arguments; God is real, and I know this as a matter of faith). The measures used are mostly just logical/maths questions.

We were following previous work that refers to the relationship between cognitive reflection and disbelief as "analytic atheism." This evidently didn't work for R1, so we have clarified it throughout. We still discuss our general reservations about the analytic atheism hypothesis in the Discussion, merely to note that approaches that make this central (e.g., cognitive byproductists describe analytic thinking as the main source of atheism) are probably mistaken.

The terminology about ‘replication-plus’ seems… unnecessary? Is this an established idea and name? It feels as though you’re rebranded the tried-and-true method of reading the literature, synthesizing it, and extending it by a small, sensible, incremental step. What am I missing here that makes ‘replication plus’ different from the ordinary process of reflection and extension? Even if this is the case, I would strongly urge the authors to use their limited word count to better justify their choice of measures, or the clarity of their analysis.

We agree that the read-and-incrementally-advance approach is longstanding and sensible. Indeed, the whole point of our ‘replication-plus’ discussion was to note that in its zeal for reform and rigor, a lot of the current replication work treats very narrow tests on single operationalizations as a favored approach. We think this narrow approach has both limited use and outsized influence.

That said, it’s clear that this intended message wasn’t transmitted clearly, so we have excised this discussion.

I generally refrain from commenting on framing, because the paper is not my own, and I believe the authors should have the right to whatever frame they want (and the reader can ignore those things anyway, paying attention to the data and analysis alone if they so wish). That said, I raise the following cautiously, and I mean no offence, but I cannot see a clear justification for these decisions.

The authors discuss both ‘veories and WEIRD psychology, lauding the value of formal theories and non-WIERD samples. Yes this research is both informal (with typical elements of statistical inference) and WEIRD. In the discussion the authors write: If this general pattern holds across societies [of our findings], we predict that—beyond religion—veories developed by WEIRD researchers to explain the weird mental states of WEIRD participants can only aspire to ever more precisely answer a mere outlier of an outlier of our most important scientific questions about human nature.

Are the authors claiming their own findings, presented here, are a ‘mere outlier of an outlier’? This is the penultimate paragraph of your paper. The paper celebrates non-WEIRD research but concludes with a call for other scholars to do the harder parts of the research? The impression I received when reading this - which I fully acknowledge wouldn’t have been the wilful intention of the authors - is to associate the virtues of formal, and non-WEIRD, on their own work which is is neither.

We have taken more efforts to contextualize the need for more beyond-WEIRD research. Representative samples like ours can help somewhat (indeed, the original WEIRD paper explicitly contrasts student samples from the general US population). But they’re just one small step. We still include some calls for de-WEIRDing our science. But we’ve toned it down, and no longer juxtapose formal and informal theories. That’s a topic for another paper. Regarding veories versus formalized theories, the original version of the manuscript never claimed that our paper was a formal theory: we merely noted that the predictor derived from a formal theory outperformed the predictions that came from prominent

veories in the field. We find this interesting. But for clarity and brevity we've dropped the veory-vs-formal-model discussion.

Minor points:

The SBS paper is cited as Jong 2012, but that's the 'fox hole' paper. The Jong et al paper with the scale is 2012, I believe.

Good catch

The authors mask themselves on the title page, but their links to their OSF in the MS include Will Gervais' name in the MS, and the OSF files list all other authors. It is the case I already knew who the authors on this MS were prior to accepting the review (and declared this with the editor). So I suppose I raise this for future reference (the authors may look into toggling how to blind their osf profiles for such materials, and using a link-masking service for the addresses).

Between a prior preprint and also the topic of the paper, I didn't think full blinding was especially feasible on this paper anyways. We masked what we could.

Concerns about the pre-reg
I commend the clarity of the pre-reg document. However, there are some incongruencies.

I.1 says that linear and quadratic terms will be analysed. Assuming the individual replication analysis are bayesian regressions, this has not been done. 'Religion' for I.1, I.2, I.3, and I.4 is not clearly specified (i.e., jong vs. binary measure. Please clarify).

This departure is explained in a footnote of the main document: "We preregistered a possible quadratic relationship between mentalizing and disbelief. For theoretical and statistical reasons, we depart from preregistration and don't analyze the quadratic here. See online Supplement for further discussion."

In the Supplement, we note that the quadratic test was nonsensical and we realized this before testing any models: "After the preregistration but before data collection and analysis, we realized that the polynomial approach was a very poor test of this idea and invites model overfitting among other ills (McElreath, 2016). The preregistered models including a quadratic for mentalizing were theoretically dubious and statistically naive, so we left them out of main analyses. We checked a few of the primary models to see if inclusion of a quadratic did much. It did not. Information criteria (WAIC) suggested that models were always better without a quadratic term for mentalizing, and the quadratic term itself never predicted much. Additional exploration about a possible low-end mentalizing blip in atheism may warrant future research with a statistically appropriate model."

The authors appear to have entirely skipped several hypotheses pertaining to the interactions. If this paper is the paper the pre-reg is associated with, these need to be run and reported.

Again, this is discussed in both a footnote and the Online Supplement: "Preregistered analyses probing for interactions with mentalizing yielded nothing of particular note and are summarized in the Online Supplement."

The authors also talk about machine learning and split-half cross-validation techniques. These are not done. Is there another paper in which they are being reported?

Were exploratory analysis conducted? If so, they should at least be acknowledge in the paper (e.g., We conducted additional exploratory analysis but do not report them here). I know that if this were my data, I would have explored it, and I would have included as a predictor in the full model religious-participation. Since the intention to conduct exploratory analysis was registered, would the authors be willing to run their full model with the inclusion of the religious-participation variable? If not exploratory analysis were conducted, this should also be acknowledged (since the intention to do so was declared).

A final point about the pre-reg - to the best of my knowledge, and I'm afraid I dn't have a convenient citation - split-half datasets with exploration and confirmation are not demonstrations of validity, but only reliability. I hope this is kept in mind when the technique is employed. It is also the case that the use [of] machine learning techniques to squeeze even more predictive power out of our dataset and we will use machine learning to suck out even more exploratory predictive power does not constitue a pre-registered analysis (though of course, such details of training will be discussed if such analyses are reported elsewhere).

We have not (yet) done any of the exploratory work or machine learning stuff. When we wrote the preregistration years ago, we wanted to note that at some point we might do this work IN ADDITION to the focal analyses. These are ancillary projects that we might yet pursue. But they are not central to the present paper. At this stage, we really just want to get our primary analyses out there and release our data so that others can explore them in whatever ways they find theoretically relevant. It's a big, expensive dataset and we suspect that many eyes and minds will find more of use in it than we have so far. We noted potential exploration and machine learning in the prereg for transparency and posterity.

I enjoyed this paper, and believe it makes a meaningful contribution to the literature. However, there are several issues the authors need to address in some depth before I can recommend publication. Good luck. I am happy to read this paper again.

Thank you for the generally positive appraisal, and also for highlighting some areas where our prose was unclear. The revision is now (we hope) much clearer about what inferences we can and cannot make.

Reviewer: 2

Comments to the Author
This is a theoretically elegant and exceptionally well written paper. This paper will be of interest

to a wide range of social science scholars and is highly consequential for evolutionary theory of religion. A few minor questions and comments for further consideration.

Thank you! We hope this paper can settle some debates in the religion literature and also contribute to broader cultural and evolutionary work and discussions.

1) The supernatural belief scale used in this study represents only a time fraction of supernatural beliefs. It is probably fine for a Judeo-Christian belief in God scale, but I think it is worth commenting more about the limitations (from a generalizability perspective across faiths and cultures). I appreciate the care that went into the measures chosen (all theoretically relevant) but wish a more wide-ranging measure of supernatural belief was included (to better capture spiritual, non-religious folks).

2) Belief in God (or gods) is a tiny fraction of the supernatural belief spectrum. What would you predict (and what are the implications for the theories tested in this paper) for spiritual/supernatural beliefs more broadly?

We now discuss these two points as an explicit limitation (p. 20): "We measured and tested predictors of religious belief and disbelief. This outcome measure is quite narrow in scope, in terms of the broader construct of religiosity. Further, our supernatural belief scale -- while it has been used across cultures -- is fairly Judeo-Christian-centric. We suspect that a broader consideration of religiosity in diverse societies may yield different patterns. The WEIRD people problem isn't just a sampling issue; it also reflects an overreliance on the theories and constructs developed by WEIRD researchers to test their weird hunches."

3) The authors mention the need for non-WEIRD (non-US) data in passing, but I think this is a major limitation of this dataset and should be discussed in the context of theoretical implications. In some ways the US is the least representative country in the world for studying atheism (which surely the authors would acknowledge).

We explicitly discuss the need for more diverse samples on this topic in the General Discussion. The US is a viable locale for testing the hypotheses and theoretical models we compare in this paper in large part because some of these models emerged based on work in the USA and they make explicit predictions about the prevalence of atheism in, say, the USA compared to Western Europe. The USA is simultaneously a subpar site for studying atheists in general (there aren't many), while also being a great place for testing theories about atheism.

Overall, we thank the editor and reviewers for their comments. They have helped us to 1) reframe our Intro to make the work more accessible to researchers inside and outside of the evolution of religion and atheism world, 2) more fully contextualize our results and their limitations, 3) clarify our analyses and their robustness, 4) explain the rationale behind our analyses, and 5) situate our work within broader literatures.