Cognition 187 (2019) 156–166

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Original Articles

The need for intellectual diversity in psychological science: Our own studies of actively open-minded thinking as a case study

Keith E. Stanovich[a], Maggie E. Toplak[b,*]

[a] University of Toronto, Canada
[b] York University, Canada

ARTICLE INFO

Keywords:
Actively open-minded thinking
Religiosity
Myside bias

ABSTRACT

Actively open-minded thinking (AOT) is measured by questionnaire items that tap the willingness to consider alternative opinions, the sensitivity to evidence contradictory to current beliefs, the willingness to postpone closure, and reflective thought. AOT has been found to be a strong predictor of performance on heuristics and biases tasks and of the avoidance of reasoning traps such as superstitious thinking and belief in conspiracy theories. Recently, several studies that have employed short forms of the AOT scale have shown startlingly high negative correlations with religiosity (in the range of −0.50 to −0.70). In a re-analysis of a large dataset, we demonstrate that it was a particular type of AOT item (termed a belief revision item, BR) that accounts for these large correlations. To our consternation, we realized that it was our research team that had introduced these items into the literature two decades ago, but we had heretofore never realized the potential for these items to skew correlations. In a new experiment, we demonstrate how BR items of this type disadvantage religious-minded subjects, and we show that it is possible to construct BR items with parallel content that are not so demographically biased. We also show that unbiased BR items do not sacrifice the predictive power that has previously been shown by AOT scales. We believe this lesson in item construction resulted from the lack of intellectual diversity in our own laboratory (specifically, the overwhelmingly secular composition of our lab personnel). We believe this case study shows the importance of intellectual diversity in psychology, especially when studying such topics as religiosity and political attitudes.

In the early 1990s, Keith Stanovich and Richard West embarked on a research program in which we attempted to measure the thinking disposition of actively open-minded reasoning. We were inspired by the writings of Baron (1985, 1988) on this concept—a thinking disposition encompassing the cultivation of reflectiveness rather than impulsivity; the desire to act for good reasons; tolerance for ambiguity combined with a willingness to postpone closure; and the seeking and processing of information that disconfirms one's beliefs (as opposed to confirmation bias in evidence seeking).

The items on the initial version (Stanovich & West, 1997) of our actively open-minded thinking (AOT) scale tapped reasoning styles such as the disposition toward reflectivity using items like: "If I think longer about a problem I will be more likely to solve it," and "Intuition is the best guide in making decisions" (the latter reverse scored). Other items tapped willingness to consider evidence contradictory to beliefs (e.g., "People should always take into consideration evidence that goes against their beliefs") and the willingness to consider alternative opinions and explanations ("A person should always consider new possibilities"). Some items tapped the willingness to postpone closure ("There is nothing wrong with being undecided about many issues"). Philosophically, the original scale focused strongly on issues of epistemic self-regulation raised in philosophical discussions (Goldman, 1986; Harman, 1995; Nozick, 1993; Samuelson & Church, 2015). The scale was a marker for the avoidance of epistemological absolutism; willingness to perspective-switch; willingness to decontextualize; and the tendency to consider alternative opinions and evidence.

We have been investigating actively open-minded thinking for almost two decades now. We have continued to refine the initial scale in several subsequent studies (Sá, West, & Stanovich, 1999; Stanovich & West, 2007; Stanovich, West, & Toplak, 2016). For example, Sá et al. (1999) introduced nine new items into the scale in order to measure an aspect of AOT that we termed belief identification. These items were inspired by a theoretical paper by Cederblom (1989) in which he argued for a potential thinking style centered around the extent to which people identify their beliefs with their concept of self (e.g., "Certain beliefs are just too important to abandon no matter how good a case can

* Corresponding author at: 126 BSB, Department of Psychology, York University, 4700 Keele St., Toronto, Ontario M3J 1P3, Canada.
E-mail address: mtoplak@yorku.ca (M.E. Toplak).

https://doi.org/10.1016/j.cognition.2019.03.006
Received 4 December 2018; Received in revised form 14 February 2019; Accepted 8 March 2019
0010-0277/ © 2019 Elsevier B.V. All rights reserved.

be made against them" [reverse scored]). Other additions and subtractions of components occurred over the next decade. By 2007 (Stanovich & West, 2007) we had a 41-item instrument that was subsequently trimmed down to 30 items in our Comprehensive Assessment of Rational Thinking (Stanovich et al., 2016; 16 items in the short form).

In our initial studies (Stanovich & West, 1997) we found that our AOT scale was moderately associated with the ability to evaluate arguments. This association held even when the variance due to cognitive ability was partialled out. In several subsequent studies, we found that our AOT scale predicted performance on a variety of heuristics and biases tasks after partialling cognitive ability (Kokis, Macpherson, Toplak, West, & Stanovich, 2002; Sá & Stanovich, 2001; Sá, Kelley, Ho, & Stanovich, 2005; Stanovich & West, 1998, 2008; Stanovich, 1999). We subsequently observed this performance pattern with a plethora of heuristics and biases tasks, including: noncausal baserate tasks, hypothesis evaluation tasks, four-card selection tasks, covariation detection, gambler's fallacy, sample size problems, conjunction fallacy, Bayesian reasoning, framing problems, ratio bias, sample size problems, and probability matching (Stanovich et al., 2016; Toplak, West, & Stanovich, 2011, 2014; West, Toplak, & Stanovich, 2008). Other laboratories have found that actively open-minded thinking measures predict performance on heuristics and biases tasks and other reasoning paradigms (Baron, Scott, Fincher, & Metz, 2015; Haran, Ritov, & Mellers, 2013; Heijltjes et al., 2014, 2015; Mellers et al., 2015; Pennycook, Cheyne, Barr, Koehler, & Fugelsang, 2014).

Our efforts, plus those of Baron, who also has continued to develop the concept (Baron, 2008, 2018; Baron et al., 2015; Baron, Gürçay, & Metz, 2017), have resulted in increasing application of the AOT concept into new and diverse areas. For example, measures of AOT have been linked to the avoidance of paranormal and supernatural beliefs (Stanovich et al., 2016; Svedholm & Lindeman, 2013). They have been linked to optimal information acquisition (Haran et al., 2013); belief in evolution (Deniz, Donnelly, & Yilmaz, 2008; Sinatra, Southerland, McConaughy, & Demastes, 2003); skeptical processing of fake news (Bronstein, Pennycook, Bear, Rand, & Cannon, 2018); accuracy in future forecasting (Mellers et al., 2015); utilitarian decision making (Baron et al., 2015); skeptical attitudes toward alternative medicine (Svedholm-Häkkinen & Lindeman, 2018); resistance to conspiracy beliefs (Stanovich et al., 2016; Swami, Voracek, Stieger, Tran, & Furnham, 2014); and cognitive inhibition skills (Campitelli & Gerrans, 2014).

1. The AOT scale encounters religion and politics

Recently, however, we have noticed some results in the literature involving AOT scales that have given us pause and have caused us to re-examine how our original scale was constructed and revised. For example, Piazza and Landy (2013) have reported some extremely high negative correlations between an AOT scale and various measures of religiosity: −0.58 with an attitudes toward religion scale, −0.59 with a religious faith questionnaire, −0.63 with a Christian orthodoxy scale, −0.58 with self-reported religiosity, and a truly astonishing correlation of −0.70 with a morality founded on divine authority scale. Baron et al. (2015) observed a correlation similar to those of Piazza and Landy (−0.61) between an AOT measure and a 4-item measure of belief that moral authority derives from God. Likewise, Bronstein et al. (2018) reported a similarly high correlation of −0.67 between a short-form AOT scale and a religious fundamentalism measure. Other studies, such as that of Yilmaz and Saribay (2017), have found similarly strong correlations (−0.47) between AOT scores and social conservativism.

We were startled by these high correlations (in the −0.50 to −0.70 range) because we have run over a dozen studies employing versions of AOT scales (see Stanovich et al., 2016) in which religiosity and political ideology have been included as demographics questions. We have very consistently found correlations in the much lower range of −0.25 to −0.40 between religiosity and the AOT (correlations with ideology are almost always even lower). The purpose of our research has not been to study religion or ideology. Instead, our goal was the development of our Comprehensive Assessment of Rational Thinking (CART) measure (see Stanovich et al., 2016). Because our focus was the CART, and not issues of religion and ideology, we have not reported these associations between religiosity and AOT until now.[1]

When puzzling over the cause of these discrepancies in the association between religiosity and AOT that our group observes versus those reported in these other studies, one of the first things that we noticed was that most of the other research tended to use short-form AOT scales—often short forms of less than 10 items. These are much smaller scales than the 41-item AOT measure that we were using 10 years ago (Stanovich & West, 2007) and the 30-item revised measure that we used in the CART.

More important than the sheer number of items of course is the specific composition of the short forms. Here, a deeper analysis of the items used across various studies revealed a potential source of the discrepancies between the results. That source appears to reside primarily in the so-called belief identification items mentioned above that were introduced into versions of our AOT scale by Sá et al. (1999). Recall that these items were designed to examine whether belief change in order to get closer to the truth is a more or less important value for the person than is retaining current beliefs (e.g., "Certain beliefs are just too important to abandon no matter how good a case can be made against them" [reverse scored]). These items clustered strongly as the second factor (termed Fact Resistance) in the factor analytic exploration of our earlier 41-item scale conducted by Svedholm-Hakkinen and Lindeman (2018).

Nine of these items were introduced into the AOT scale in our 1999 study. One item of a similar type was already in the earlier scale. Thus, the 41-item scale used in the mid-2000s by Stanovich and West (2007) had 10 of 41 items of this type (24.4%). The 30-item updated AOT subtest in our CART (Stanovich et al., 2016) had 9 belief-identification items (30% of the total). It was immediately of concern to us in perusing the short-forms used in other studies that the proportion of belief-identification items was substantially higher. Piazza and Landy (2013), in the study that obtained extremely high correlations with a host of variables measuring religiosity, used a 7-item AOT scale that contained five belief identification items (Yilmaz & Saribay, 2017, who found a high correlation with social conservatism, used the same 7-item short form). Baron et al. (2015) used an 8-item AOT short form that contained four belief identification items. Bronstein et al. (2018), who reported a substantial correlation of −0.67 with a religious fundamentalism measure, used an 8-item AOT short-form that contained five belief identification items. Thus, the three studies displaying correlations of 0.55 or above between AOT and religiosity-related variables used short-forms of the AOT that were composed of 71.4%, 50%, and 62.5%, belief identification items. This is much higher than the roughly 24%–30% composition that we have used in our versions of this instrument. Thus, our initial conjecture was that the high proportion of belief identification items in these other studies is the source of the higher correlations with religiosity that they obtain.

2. How might these items be inflating correlations with religiosity?

What is the feature of the belief identification items that might be augmenting correlations with religiosity? The following three items capture the central logic of this type of item:

Beliefs should always be revised in response to new information or evidence.

Certain beliefs are just too important to abandon no matter how

[1] Note that the AOT is a supplemental measure in the CART and not part of the total CART score.

good a case can be made against them. (reverse scored)

One should disregard evidence that conflicts with your established beliefs. (reverse scored)

The general thrust of this kind of item is that the subject is being asked what they do when they encounter evidence that conflicts with a prior belief or opinion. A willingness to alter the prior belief results in a high score on AOT items of this type. Dismissing the new evidence and leaving the prior opinion unaltered results in a low score on this type of item. In essence, these items ask the subject to directly assess their degree of belief bias or myside bias[2] (Evans, 2002, 2017; Mercier, 2017; Stanovich, West, & Toplak, 2013; Taber & Lodge, 2006)—how much they prefer conclusions consistent with what they previously believed as opposed to conclusions consistent with new evidence. For clarity in the subsequent discussion, we will call these the BR (belief revision) items on AOT scales. Note that a BR item is not simply an item that uses the *word* belief, but an item with a specific structure—one that in effect queries whether the subject alters a previously held opinion in the face of conflicting evidence.

It is important to note that no *particular* prior opinion is mentioned in any of these items. It is just the generic word "belief" that is used. Of course, adjusting a prior belief based on new contradictory evidence is more or less easy to do depending upon what the belief is. For example, is it my belief that I voted the right way in the presidential election of 2016? Or is it my belief that the deli counter is better at Albertsons than at Safeway? The latter is obviously going to be a belief that is more easily conditioned by evidence than is the former. The generic nature of the word "belief" in these items allows the respondent to insert any imaginary opinion for the belief in question. Potential social desirability considerations may lead most people to insert a belief that is easy to change. Thus, these items are almost inviting someone to fall prey to the bias blind spot—that is, thinking that others are characterized by a particular bias but that you yourself are not (Pronin, 2007).

All of the above might well be true for a secular person. But a person with strongly held religious convictions might well see the word "belief" as referring to their spiritual beliefs—a class of beliefs that are not going to be easily altered by evidence. In contrast, a secular person might be much less likely to see the word "belief" as denoting an imaginary opinion that is so strongly held. Our conjecture is that what a religious person does when seeing the generic word "belief" is to simulate an *actual* stance (their spirituality) that is much more difficult to reconsider based on evidence than a generic belief, or an anodyne one. To see this, one might imagine a secular person who answered one of our belief identification items affirmatively. To such a person who answered by saying: "Well of course I'd change my belief if I got contradictory evidence, that's what an intelligent person does" we might imagine the conversation continuing. "OK," we might reply, "now imagine your belief is your vote against Trump in 2016. Would you be likely to change that based on new information?" It is doubtful that the item would be so enthusiastically endorsed if we substituted for the generic belief slot in the question the *specific* belief "my vote against Trump was a good thing". In short, such items are much easier to score well on if one does not insert for the generic word "belief" a specific strongly held opinion (note that scoring well here means that the subject indicates they would give up the belief or change it in the face of evidence).

A more general way to think about our argument is in terms of the concept of a protected value. Protected values are those that resist trade-offs with other values, particularly economic values (Baron & Spranca, 1997). These are values that are viewed as moral obligations that arise from deontological rules concerning action, and the thought of violating them often provokes anger. Experiments have shown that subjects are reluctant to trade and/or compare items when protected values are at stake (Baron & Leshner, 2000; Bartels & Medin, 2007). Needless to say, the belief in a protected value will not be easily altered by any kind of evidence. If a subject inserts an actual protected value into the open-ended "belief slot" in a BR item, such a subject will necessarily score poorly on that item. Our conjecture is that the word belief might entice religiously-minded subjects to do just that.

In short, to the extent that secular people are inserting an anodyne belief like the preference for Pepsi over Coke as opposed to a belief strongly related to worldview like belief in God or a particular religion, then they are advantaged on such items. This advantage, and the corresponding disadvantage to the religious respondent who might slot in for the term "belief" belief in God (a protected value), inflates the negative correlation between AOT and religiosity—such items are harder to agree to on the part of the religious-minded. This non-equivalence never occurred to us at the time we were making up the belief identification items, perhaps because of our own secular biases—exacerbated by the fact that we were not specifically studying the AOT-religiosity association.

3. Testing the conjecture in data from the CART

We were able to test our theoretical conjecture with data that we already had in hand. The final study testing our rational thinking test, the CART (labelled RT60 in Stanovich et al., 2016), involved 747 subjects who completed a 30-item version of our AOT scale which contained both BR and non-BR items. The demographics questions that these subjects completed assessed both belief in God and political ideology.

We recoded the items on the 30-item AOT scale to identify the BR items. Not all belief identification items in the scale had the logic described above—one whereby a person is asked to alter or abandon a belief based on new or contradictory evidence. Likewise, some items outside of the original belief identification set had the BR logic. The resulting reclassification yielded 7 BR AOT items and 14 non-BR AOT items. We eliminated from this analysis some items that tapped belief identification but did not have the BR logic and others that otherwise less clearly tapped a non-BR AOT concept. The 14-item non-BR scale was focused on such central AOT concepts as the disposition toward reflectivity ("Coming to decisions quickly is a sign of wisdom" (reversed scored); the willingness to consider alternative opinions and explanations ("A person should always consider new possibilities"); tolerance for ambiguity combined with a willingness to postpone closure ("There is nothing wrong with being undecided about many issues"); and the so-called master rationality motive (Stanovich, 2008)—the desire to act in accordance with reasons ("I like to think that my actions are motivated by sound reasons").

Table 1 presents the correlation between the 7-item BR AOT and the 14-item non-BR AOT scale (Cronbach's alpha 0.84 and 0.79, respectively) with a selection of variables from the study. The very first pair of correlations concerns the belief in God demographic question—a seven-point scale anchored at one with "I am certain that God exists" (scored as 7) and at the other end with "I am certain that God does not exist" (scored as 1). Consistent with our conjecture above, the −0.53 correlation with the seven BR AOT items was significantly larger than the −0.25 correlation with the fourteen non-BR AOT items using the Steiger (1980) test for dependent correlations ($t(744) = 9.92$, $p < 0.001$). The difference in correlations was not driven by just one or

[2] The terms confirmation bias, belief bias, and myside bias are used in a confusing manner in the literature. Myside bias has sometimes been viewed as a subclass of confirmation bias (Hahn & Harris, 2014; McKenzie, 2004), but a confirmation bias is not necessarily a myside bias (Mercier, 2017), something that has been known for quite some time (Klayman & Ha, 1987). For our purposes here, confirmation bias refers to the tendency to confirm the focal hypothesis (*however* the hypothesis is made focal). Belief bias is processing with a bias toward current beliefs or hypotheses. Myside bias refers to processing in favor of currently highly-valued existing opinions. It refers to beliefs that we want to be true. In short, people display myside bias when they evaluate evidence, generate evidence, and test hypotheses in a manner biased toward their own opinions (Stanovich et al., 2013).

Table 1
Comparison of correlations involving the BR-AOT items versus the non-BR AOT items.

	7 Belief Revision AOT Items	14 Non-Belief Revision AOT Items
Study RT60 (all subjects: N = 747)		
Belief in God	−0.53	−0.25
Liberal ideology	0.37	0.22
Superstitious thinking	−0.53	−0.49
Belief in conspiracy theories	−0.33	−0.38
Study RT60 (mechanical turk sample: N = 397)		
Belief in God	−0.43	−0.17
Liberal ideology	0.33	0.20
Superstitious thinking	−0.57	−0.51
Belief in conspiracy theories	−0.40	−0.42
Study RT60 (university sample: N = 350)		
Belief in God	−0.47	−0.23
Liberal ideology	0.31	0.17
Superstitious thinking	−0.31	−0.39
Belief in conspiracy theories	−0.16	−0.28

All correlations significant at the 0.001 level.

two outlier items. In fact, it was very consistent across items. The seven correlations between individual BR items and belief in God ranged from 0.29 to 0.51, whereas the 14 correlations between individual non-BR AOT items and belief in God ranged from 0.07 to 0.22. That, is, not a single non-BR item correlated as highly with belief in God as did the least strongly-correlating BR item.

The next row of the table indicates that the correlation with liberal ideology—a six-point scale ranging from very conservative (scored 1) to very liberal (scored 6) was also significantly higher for the 7-item BR AOT subscale than for the 14-item non-BR AOT subscale (0.37 versus 0.22; test for dependent correlations: $t(744) = 4.78$, $p < 0.001$). The next two rows of the table correlate the two AOT scales with a measure of superstitious thinking (12 items) and the measure of belief in conspiracy theories (24 items) described in Stanovich et al. (2016) and show that the two AOT subscales are both moderately good predictors of a similar magnitude. The overall conclusion here then is that the seven BR AOT items and the fourteen non-BR AOT items are just as strongly related to certain criterion variables such as superstitious thinking and belief in conspiracies, but the former correlates much more strongly with religiosity and political ideology.

The 747 subjects in this particular study, RT60, were actually comprised of two different groups. One group of 350 subjects were university students recruited on the James Madison University campus. The other group of 397 subjects were recruited using the Amazon Mechanical Turk and thus were very different in demographics from the university sample, as has been much discussed (Paolacci & Chandler, 2014). For example, the mean age of the MTurk group in our sample was over 12 years older than the university group. The MTurk group was majority male (58%), whereas the university sample was majority female (69%). The remaining correlations in Table 1 break out the results for the two subsamples in order to ascertain whether the overall trends discussed above held in both. In fact, the trends were consistent in each sample. In both subsamples, religiosity was more highly correlated with the BR AOT items than with the non-BR AOT items (−0.43 versus −0.17, and −0.47 versus −0.23). In both subsamples, ideology was more strongly related to the BR AOT items than to the non-BR AOT items. Finally, the non-BR AOT items were just as strongly predictive of superstitious thinking and belief in conspiracy theories as were the BR AOT items. In neither sample does one give up predictive accuracy by using AOT items that are less strongly related to religiosity and ideology.

In summary, the results indicate that there is something specific about the belief revision AOT items that leads to higher correlations with religiosity. The patterns displayed in Table 1 are consistent with our conjecture that the feature of these items that leads to the higher correlations is the word “belief” and its tendency to focus a religious subject on the specific content of spiritual belief while leaving the secular subject free to insert as content virtually anything they wish. It is much easier to give a high AOT response to these items if the actual belief inserted as imaginary content is anodyne rather than something that reflects a deep commitment or a protected value. This conjecture can be straightforwardly tested. The test would consist of varying the content of the issue in what we might call the belief slot that these items create. That is, in addition to testing the generic word belief as is normally the case in these AOT items, a specific instruction to insert religious belief could be given. Finally, another condition might be created which inserted content about which both religious *and* secular participants might hold a strong prior opinion. This is, in essence, the experiment that we conducted.

4. An experimental test: Actively open-minded thinking is harder when the belief is real

The subjects in our experiment completed an 18-item AOT scale. Half of the scale (9 items) was composed of non-BR items (e.g., “A person should always consider new possibilities”) and these were completed by all of the subjects (643 in total). The other nine items were belief revision AOT items. There were three different kinds of BR items created for this study and subjects were randomly assigned to receive one of the three types. One type we have discussed previously—items that use the generic word “belief” without having content specified for the belief. These will be termed the Generic BR items. Three examples of this type of item are given in Table 2. One third of the subjects (n = 214) were randomly assigned to the Generic BR condition.

Another third of the sample (n = 213) was randomly assigned to what we will term the Religious BR condition. Here, the word “belief” is replaced by wording that makes it clear that the item is referring to religious belief. Three examples of this type of item are given in Table 2. Finally, one third of the sample (n = 216) was randomly assigned to a third condition that we will term the Secular BR condition, where inserted for the word “belief” was a specific opinion on a secular political issue. Note that in the Secular condition the items do not specify a

Table 2
Examples of the items used in the three belief revision conditions.

Generic Condition Example #1	People should always take into consideration evidence that goes against their beliefs.
Religious Condition Example #1	People should always take into consideration evidence that goes against their religious beliefs.
Secular Condition Example #1	People should always take into consideration evidence that goes against their beliefs about capitalism.
Generic Condition Example #2	Beliefs should always be revised in response to new information or evidence.
Religious Condition Example #2	Religious beliefs should always be revised in response to new information or evidence.
Secular Condition Example #2	My opinions about President Trump should always be revised in response to new information or evidence.
Generic Condition Example #3	Certain beliefs are just too important to abandon no matter how good a case can be made against them. (R)
Religious Condition Example #3	Certain religious beliefs are just too important to abandon no matter how good a case can be made against them. (R)
Secular Condition Example #3	Certain beliefs about inequality and fairness are just too important to abandon no matter how good a case can be made against them. (R)

R = reverse scored.

Table 3
Means for the three conditions across the two AOT subtests and two other variables.

	Condition			
	Generic	Religious	Secular	F(2.640)
Belief revision AOT items	42.3	40.9	37.9	17.63***
Regular non-BR AOT items	43.7	43.8	42.7	3.63*
Superstitious thinking	21.1	21.6	22.2	0.58
Belief in conspiracy theories	35.0	35.7	36.8	1.06

* $p < 0.05$.
*** $p < 0.001$.

particular valence for the belief (positive or negative). Table 2 contains three examples of this type of item[3]. Each of the three groups of subjects received nine BR items that were all the same type (Generic, Religious, or Secular; Cronbach's alpha = 0.85, 0.87, and 0.75, respectively). All of the subjects received the same non-BR AOT items (Cronbach's alpha = 0.66).

The 18 AOT items were intermixed with the items from two other scales. Subjects completed an 11-item superstitious thinking scale (Cronbach's alpha = 0.91) based on the parallel subtest in the CART that we have used (Stanovich et al., 2016) and on other research (sample item: "I have personal possessions that bring me luck at times"). Subjects also completed a 12-item conspiracy beliefs scale (e.g., "High-level U.S. government operatives knew ahead of time that the 9/11 attack on the World Trade Center was about to occur"), a shortened version of that subscale from the CART (Cronbach's alpha = 0.90). For each of the scales, subjects responded on a six-point scale with no neutral point: disagree strongly (1), disagree moderately (2), disagree slightly (3), agree slightly (4), agree moderately (5), agree strongly (6). Intermixed were also three impression management items and 12 items of a pilot study of another scale.

After completing the three intermixed questionnaires, the subjects answered a series of demographic questions that included two questions on their religious beliefs. The first was "Religion is important in my everyday life" and it was answered on a six-point scale ranging from disagree strongly (scored as 1) to agree strongly (scored as 6). The second question was "My feelings concerning the existence of God are" and was answered on a seven-point scale ranging from "I am certain that God does not exist" (scored as 1) to "I am certain that God exists" (scored as 7). The responses on these two questions were standardized and summed to yield the religiosity score. We acknowledge that the God question might be less applicable to non-monotheistic religions and might need to be revised for cross-cultural work and samples not restricted to United States citizens as ours was.

The subjects answered two demographic questions regarding their political ideology. The first was "*Economically*, I would consider myself to be" and was answered on a six-point scale ranging from very conservative (scored 1) to very liberal (scored 6). The second was "*Socially*, I would consider myself to be" and was answered on a six-point scale ranging from very conservative (scored 1) to very liberal (scored 6). The responses on these two questions were standardized and summed to yield the liberal ideology score.

The experiment was run using Qualtrics software. Subjects (N = 643) were recruited from Prolific (Palan & Schitter, 2018; Peer, Brandimarte, Samat, & Acquisti, 2017). They all had English as a first language and all were United States citizens. Their mean age was 33.7 years (SD = 12.3) and 50.9% were female. At the close of the experiment, 214 had been assigned to the Generic condition, 213 had been assigned to the Religious condition, and 216 had been assigned to the Secular condition.

Table 3 presents the means for each of the four variables (superstitious and conspiratorial thinking scales plus the BR and non-BR subscales of the AOT). The last two rows indicate that, as expected, the three randomized groups performed similarly on the superstitious thinking measure and the belief in conspiracy theories scale. This is expected because the three randomized groups each received the same version of both of these scales. The table does indicate however that there was a small difference on the regular non-BR AOT items. This was not expected, since these items were the same across all three groups. The difference was small however ($\eta^2 = 0.011$) and was significant because of the relatively large sample size. The nature of the difference was that the mean of the Secular group was about one point lower than that in the other two groups. This could simply be a random difference, or perhaps it is due to the Secular BR items somehow priming and reducing performance on the non-BR items.

Most important however, is the first line of the table which indicates that there were significant differences among the three randomized groups on the belief revision AOT items ($p < 0.001$, $\eta^2 = 0.052$). Scheffe post-hoc tests indicated that the mean of the Secular group was significantly smaller than either of the other two. The difference between the Generic and Religious means did not reach significance. The direction of this difference was as expected, given our theoretical conjecture. The Generic condition allows *any* imaginary belief to be inserted for the ambiguous term belief, thereby almost inviting the subjects to imagine a non-entrenched belief that makes high-AOT responding easier. A particular belief is focal in the Religious condition (a belief that might be hard to condition based on evidence for some subjects) and this makes it harder to score well on such AOT items, at least for the religiously-inclined. The Secular condition makes particular content focal for *all* subjects—religious and secular. This prevents a subject from imagining an anodyne belief and then scoring highly on such an item by indicating flexibility on the response scale. Instead, the subject is forced to indicate how they would alter beliefs that may be of some concern to them in the face of evidence. This is harder to do when faced with a particular political/cultural belief than it is in the abstract (as it is in the Generic condition).

That a response pattern like this is occurring is further indicated by the results presented in Table 4 which displays the mean performance in the three conditions for subjects high and low in religiosity in our sample. To be classified as high in religiosity, subjects had to indicate that they agreed at least slightly with the "Religion is important in my everyday life" question *and* they had to indicate at least that there "probably is a God." To be classified as low in religiosity, subjects had to indicate that they disagreed at least slightly with the "Religion is

[3] With respect to the Secular condition, the comparisons in this study are clearly quasi-experimental rather than strictly experimentally controlled. That is, it was easy to change most Generic items into Religious condition items by substituting the phrase religious belief in place of the word belief in each Generic item. However, it was clearly not possible to create the Secular condition in this manner because there is not a particular phrase that represents the secular stance. We decided to use a selection of political issues in the hope that, across a range of such issues, at least some of them would represent protected values for both religious and secular people, making the scale more balanced across these two different groups. Our strategy was to try to sample specific political issues that are under discussion currently (President Trump, capitalism, inequality, etc.). Another strategy might be to use the term "political beliefs" uniformly across items. This strategy is worth trying, but it has its own drawbacks. For example, it is possible that many people do not consider politics to be very important, and may not even care that much about elections. However, these same people may in fact care quite deeply about particular issues that a social scientist would consider clearly political, such as equal pay, rights for gay individuals, immigration, or the nature of policing. Such a person might not see the generic term "political beliefs" as representing a protected value, but may well hold specific political issues (e.g., immigration) as protected values. Nevertheless, we clearly do acknowledge the lack of strict experimental equivalence across the three conditions caused by the nature of the Secular condition. That particular condition represents more of an experimental demonstration, or perhaps an existence proof that, if subjects are blocked from inserting religious belief as the focus of such items, then the correlation of the AOT with religiosity is much reduced.

Table 4
Means in the three belief revision conditions as a function of religiosity.

	Condition		
	Generic	Religious	Secular
Low religiosity	44.7 (n = 123)	45.3 (n = 113)	39.7 (n = 108)
High religiosity	37.8 (n = 62)	33.4 (n = 56)	36.3 (n = 60)

important in my everyday life" question *and* they had to be at least "not sure whether God exists or not" or show even greater skepticism. These joint criteria classified 198 subjects as high in religiosity, and 344 subjects as low in religiosity (101 subjects were not classifiable).

Overall, a 2×3 analysis of variance (low/high religiosity by Generic/Religious/Secular condition) indicated a significant tendency for the subjects low in religiosity to score higher on the belief revision AOT items ($F(2,536) = 139.65$, $p < 0.001$). More important is that the interaction displayed in Table 4 is statistically significant ($F(2,536) = 15.58$, $p < 0.001$), and it is in the direction supporting our unease about the Generic items commonly used in AOT scales. For the subjects low in religiosity, performance is no worse at all in the Religious condition. For them, a religious belief presents no particular difficulty—it is easy for them to imagine evidence conditioning religious beliefs because the evidence in their case would not be threatening a strongly held position. However, for these subjects, performance drops off markedly in the Secular condition—which involves beliefs and opinions that they may feel strongly about. The pattern was different for the subjects high in religiosity. Their performance dropped off in the Religious condition compared to the Generic condition. The Religious condition makes it clear that religious beliefs were focal to the statement they were evaluating. These subjects found the statements in the Secular condition easier to endorse.

Focusing on just the Religious and Secular conditions, we see a complete crossover—subjects low in religiosity found the Secular BR items harder to endorse than the Religious items; whereas the subjects high in religiosity found the Religious items harder to endorse than the Secular ones. Another way to see the interaction is to note that the subjects differing in religiosity were maximally different on the religious items (approximately 12 points difference), were minimally different in the Secular condition (about 3.5 points different), and displayed an intermediate difference (of about 7 points) in the Generic condition. Finally, the Generic condition was much easier for those low in religiosity to score higher on than the Secular condition (44.7 vs. 39.7), but the Generic condition was not much easier for those high in religiosity than was the Secular condition (37.8 vs. 36.3).

We can see how these differences between the different types of BR AOT items would affect correlations observed in various studies by examining the correlations presented in Table 5. This table presents the associations with the two AOT subscales across each of the three conditions. At the top of the table are the correlations from the 214 subjects who were randomly assigned to the Generic condition. The findings here replicate the patterns that were observed in study RT60 of the development of the CART (Stanovich et al., 2016) and presented in Table 1. Specifically, the correlation between the BR items and religiosity (−0.46) was significantly larger than the correlation between the regular non-BR AOT items and religiosity (−0.25; $t(211) = 4.13$, $p < 0.001$). The difference in correlations was not driven by just one or two outlier items. The nine correlations between individual BR items and religiosity had a median correlation of −0.29, and 8 of 9 were greater than 0.20 in absolute magnitude, whereas the nine correlations between individual non-BR AOT items and religiosity had a median correlation of −0.15, and only 3 of the 9 were greater than 0.20 in absolute magnitude.

The next row indicates that the correlation with liberal ideology was also higher for the BR items but the difference here was less so (0.37 vs 0.28) and did not quite reach significance ($t(211) = 1.74$). The ability of both types of AOT items to predict superstitious thinking and belief in conspiracy theories was similar, as indicated in the last two rows.

Table 5
Correlations across the two AOT subscales in the three experimental conditions.

Generic condition (n = 214)		
	Generic BR items	Regular non-BR AOT items
Religiosity	−0.46	−0.25
Liberal ideology	0.37	0.28
Superstitious thinking	−0.45	−0.39
Belief in conspiracy theories	−0.24	−0.19
Religious condition (n = 213)		
	Religious BR items	Regular non-BR AOT items
Religiosity	−0.60	−0.09
Liberal ideology	0.48	0.26
Superstitious thinking	−0.43	−0.46
Belief in conspiracy theories	−0.16	−0.24
Secular condition (n = 216)		
	Secular BR items	Regular non-BR AOT items
Religiosity	−0.27	−0.22
Liberal ideology	0.23	0.25
Superstitious thinking	−0.53	−0.53
Belief in conspiracy theories	−0.27	−0.29

The next set of correlations in Table 5 present the results from the Religious condition. Here, there was a massive difference in the correlation between religiosity and the BR AOT items versus the non-BR items: −0.60 versus −0.09 ($t(210) = 9.15$, $p < 0.001$). There was also a fairly large difference in the correlations with liberal ideology, 0.48 versus 0.26, that was statistically significant ($t(210) = 3.42$, $p < 0.001$). As in the Generic condition, both types of AOT items had similar correlations with superstitious thinking and with belief in conspiracy theories.

The final set of correlations in Table 5 present the results from the Secular condition. Here, the correlation between religiosity and the Secular BR items was markedly reduced, to just −0.27—not significantly different from the correlation with the non-BR AOT items, −0.22 ($t(213) = 0.91$). Likewise, the correlation with liberal ideology was similar across the two different types of AOT items[4]. Finally, both types of items had similar correlations with superstitious thinking and with belief in conspiracy theories.[5]

5. Recommendations for AOT scales

These results indicate that the use of the belief revision AOT items may yield misleadingly high estimates of the association of actively open-minded thinking and religiosity (and to a much lesser extent perhaps political ideology as well). Our results indicate that it is not necessary to use such items in order to have an AOT scale that is predictive of the types of outcome variables studied in previous research. Tables 1 and 5 indicate that AOT versions without belief revision items predict superstitious thinking and belief in conspiracy theories almost as accurately as versions with these items.[6]

[4] Consistent with other research (Baron, 2015; Carl, 2014; Pennycook & Rand, 2019; Yilmaz & Saribay, 2016), we found that the correlations with social liberalism (0.31, 0.27, 0.30) were consistently higher than those with economic liberalism (0.20, 0.22, 0.17).

[5] The correlation between the BR AOT items and the non-BR AOT items varied across the three conditions, from a high of 0.66 in the Generic condition to a low of 0.47 in the Religious condition. The two types of items displayed a correlation of 0.59 in the Secular condition.

[6] Indeed, there is an AOT item type that is even more independent of religiosity (and ideology) that predicts superstitious thinking and conspiracy theory belief almost as well—the aforementioned master rationality motive items (Stanovich, 2008). It is an item type that taps the desire to act in accordance with reasons (e.g., "I like to think that my actions are motivated by sound reasons").

Nevertheless, we acknowledge that the belief revision items represent a very valid and central component of the AOT construct because they indicate how the subject conditions beliefs when presented with evidence. If belief revision items of this type are deemed necessary, then we would recommend that items of the type used for the Secular form of the scale be used. Unlike the Religious condition which disadvantages those high in religiosity, the Secular condition does not correspondingly disadvantage secular or nonreligious individuals. Although the Secular condition employs beliefs with content, the attitude toward the content is not tipped in one direction or another. In Table 2, Example #2, the item does *not* say "negative opinion of Trump", or "my vote against Trump". It simply centers attention on opinions about a polarizing figure—but those opinions might be positive or negative. Unlike the items in the Generic condition, the Secular items put the *same* issue in people's minds, but the issue is not stated so that it is more salient to the religious or nonreligious—both of whom might have strong opinions about Trump, one way or the other.

We agree that assessing whether opinions and beliefs are flexibly conditioned by evidence is a valid component of actively open-minded thinking. However, using items without content (employing the neutral/generic word "belief") introduces unnecessarily large correlations with religiosity into the scale—unnecessary because we have shown that Secular BR items have correlations with superstitious thinking and belief in conspiracy theories that are similar to those obtained by using the Generic BR items (and also similar to those shown by the non-BR AOT items).

Unlike the Generic BR items, the Secular BR items display correlations with religiosity in the range similar to that displayed by non-BR AOT items (-0.20 to -0.30). It is not necessary to use items that inflate correlations with religiosity (or ideology) in order to predict superstitious thinking or conspiracy theory thinking with an AOT scale. As we argued previously, Generic belief revision items are difficult to the extent that the subject inserts a protected value into their imagination as an exemplar, and religious subjects might be especially prone to interpret the word belief as referring to a religious belief. In contrast, a non-religious subject might do *well* on the Generic items precisely *because* they have a large bias blind spot (Pronin, 2007) and are not prone to imagine a protected value in the "belief slot". This will be especially true if they can discern the subtle social desirability in a research scientist asking them, essentially, "How do *you* respond to evidence?" After filling in an anodyne opinion for the word "belief" it is easy for them to say, like in the aphorism, "I change my mind, of course".

It is important to note that it is only the correlations with religiosity (and to a lesser extent, ideology) that are skewed by the use of generic BR items. The correlations with other marker variables are not affected. For example, in Table 1, presenting the results from study RT60, the BR AOT items displayed correlations with superstitious thinking and belief in conspiracy theories that were very similar to those obtained with the non-BR items. The results of our new study, displayed in Table 5, are highly convergent. If we consider the correlations with the non-BR items to be the baseline, we see that all three types of BR item (Generic, Religious, and Secular) produced correlations with superstitious thinking and belief in conspiracy theories that were very similar to the baseline non-BR correlations.

The findings regarding the Secular BR items represent our most critical demonstration. The correlational relationships displayed in Table 5 represent an existence proof that it is possible to construct BR items that show correlations with criterion measures (e.g., superstitious thinking, belief in conspiracy theories) that are just as large as those obtained with non-BR items and that do not inflate the correlation with religiosity beyond what is found with the non-BR items.

6. Why is the AOT so potent a predictor of rational thinking?

It is important to measure AOT carefully and without bias for two reasons. First, as noted above, the concept is increasingly being used in areas of sociocultural concern such as: belief in alternative medical practices; belief in pseudoscience and conspiracies; detection of fake news; moral decision-making; future forecasting; and in debates about the origins and correlates of political ideologies (Baron, 2018; Bronstein et al., 2018; Stenhouse et al., 2018). Secondly, it is becoming increasingly clear that in the domain of rational thinking, AOT is a potent predictor. It is ubiquitously linked to subtests in our CART test (Stanovich et al., 2016). The CART is a very comprehensive measure of rational thought and is composed of 20 different subtests (and four supplemental scales, which include AOT). Our 30-item AOT scale is not only correlated with every one of the 20 subtests, but it accounts for variance over and above cognitive ability on the vast majority of them (17 out of 20). Despite the multifariousness of the rationality construct itself (which is why the CART contains 20 subtests), a particular thinking style—actively open-minded thinking—does permeate almost all of the components (from probabilistic reasoning to avoiding overconfidence, and many more).

What are the central features of thinking that make the AOT such a good predictor of rational thinking? We would argue that the common psychological dimension is the tendency to engage in cognitive decoupling (Stanovich & Toplak, 2012; Stanovich, 2011). To a lesser extent the items may tap a related tendency toward the decontextualization of problems. These psychological characteristics are particularly applicable to the heuristics and biases tasks that operationally define rationality in cognitive science (Baron, 2008; Kelman, 2011; Stanovich, 1999, 2012).

Cognitive decoupling is particularly important in performance on these tasks because heuristics and biases tasks often create hostile problem-solving environments. Many such tasks have an intuitively compelling wrong answer that must be overridden. Frederick (2005) famous bat-and-ball problem from his Cognitive Reflection Test (CRT) is the most well-known (A bat and a ball cost $1.10 in total. The bat costs $1 more than the ball. How much does the ball cost?). The tempting intuitive but wrong response is a feature of these problems, not a bug. In fact, the presence of the compelling intuitive response is precisely what makes the problem diagnostic of the propensity toward reflective thinking. Interestingly, some early research showed that CRT performance was negatively related to strength of religious belief (Gervais & Norenzayan, 2012). However, in a cross-cultural investigation of 13 different diverse societies, Gervais et al. (2018) found that the relationship was quite variable and that it was only robust in 4–5 of the cultures that tended to have high national levels of religiosity. They speculated that "without some sufficient level of cultural support, there may be no need for people to analytically override religious impulses" (p. 273).

In contrast, a typical heuristics and biases task is designed to trap the cognitive miser (Kahneman, 2011; Stanovich, 2004, 2018). On IQ tests, people are not tempted to engage in so-called miserly processing due to the presence of an intuitively compelling alternative. The fact that many heuristics and biases tasks can be construed by the subject in different ways (a statistical interpretation versus a narrative interpretation, for instance) is often seen as a weakness of such tasks when in fact it is the design feature that makes the task diagnostic. In a probabilistic reasoning task from this literature, the entire point is to see how dominant or nondominant the statistical interpretation is over the narrative interpretation.

Heuristics and biases tasks create a more hostile reasoning environment than typical IQ test problems in that the latter do not contain enticing lures toward an incorrect response. Neither is the construal of an intelligence test item left up to the subject. Instead, the instructions to an IQ test item attempt to remove ambiguity in a way that is not true of a heuristics and biases problem. The famous Linda conjunction problem would be a prime case in point. The instructions purposefully do not tell the subject how to weight the conflicting cues—the similarity of the description ("deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear

demonstrations") to the classification ("feminist bank teller) and subset/superset relationship between feminist bank teller and bank teller. Most subjects detect the two conflicting cues in the problem (De Neys, 2014), but the instructions pragmatically obscure the fact that the correct weighting is 0/100.

Rational thinking paradigms attempt to measure the *propensity* to use a cognitive skill in a way that IQ tests do not. A typical heuristics and biases task will pit a statistical way of viewing a problem against a non-statistical way of viewing a problem—or a control-group way of thinking against correlational thinking—in order to see which kind of thinking dominates in the situation. People who can answer an explicit probability question on a test, or who can accurately define "control-group" when asked, may not show the sensitivity to invoke these principles when their relevance to a problem is partially disguised.

Cognitive decoupling is implicated in such hostile task environments in two ways (Oaksford & Chater, 2012; Stanovich & Toplak, 2012; Stanovich, 2011). It is implicated in the inhibitory override of the intuitive response triggered by many heuristics and biases tasks, but it is also implicated in the sustained simulation of alternative worlds that is necessary to compute the correct response. The first type of cognitive decoupling—inhibition of the prepotent response—is akin to that studied in the executive functioning literature (Kovacs & Conway, 2016; Miyake & Friedman, 2012; Nigg, 2017). But the ability to suppress miserly processing gets the job only half done. Suppressing one response is not helpful unless there is a better response available to substitute for it. Where do these better responses come from? One answer is that they can come from processes of hypothetical reasoning and cognitive simulation (Evans, 2007, 2010; Evans & Stanovich, 2013). When we reason hypothetically, we create temporary models of the world and test out actions (or alternative causes) in that simulated world. In order to reason hypothetically we must, however, have one critical cognitive capability that is a type of cognitive decoupling: we must be able to prevent our representations of the real world from becoming confused with representations of imaginary situations.

To engage in exercises of hypotheticality and high-level cognitive control, one has to explicitly represent a psychological attitude toward the state of affairs as well as the state of affairs itself (Dienes & Perner, 1999; Evans & Over, 1999). Thus, decoupled representations of actions about to be taken become representations of potential actions, but the latter must not infect the former while the mental simulation is being carried out. Dealing with these so-called secondary representations—keeping them decoupled—is costly in terms of cognitive capacity. The tendency to initiate such decoupling for the purposes of simulation is a dispositional variable, separable from cognitive capacity (the ability to desist and to sustain the decoupling). It is particularly important in the context of heuristics and biases tasks because they are constructed in a hostile manner. They are tasks that require override of a prepotent response and also necessitate that a hypothetical representation be entertained at the same time the prepotent response is being suppressed.

Given this understanding of the importance of cognitive decoupling in heuristics and biases tasks, our conjecture is that AOT scales tap the propensity to engage in these types of cognitive operations. For example, many AOT items relate to avoiding miserly processing and overriding the tendency to fix beliefs quickly or to decide quickly: "If I think longer about a problem I will be more likely to solve it," "Coming to decisions quickly is a sign of wisdom". Others tap the willingness to consider possibilities beyond the focal model that is in the mind, e.g., "Considering too many different opinions often leads to bad decisions (reverse scored)," "Changing your mind is a sign of weakness, (reverse scored)," "A person should always consider new possibilities." Additionally, all of the belief revision items require the subject to hold an existing belief in abeyance while simulating the effect of new information on the original belief—classic cognitive decoupling.

In an earlier book, Stanovich (2004) argued that the requirements of modernity serve to increase the value of cognitive decoupling. This argument has as a corollary that there might well be cultural differences in decoupling frequency. Indeed, we earlier argued in our discussion of the cross-national findings of Gervais et al. (2018) regarding the CRT that the propensity to decouple might interact with culture. To the extent that the AOT taps this propensity, its relationships might show cross-cultural variation. To date, little cross-cultural work has been done with the AOT, but relationships found in North America do at least appear to replicate in European settings (e.g., Svedholm-Häkkinen & Lindeman, 2018).

7. A case study of ideological bias in psychological science—our own!

We view the rethinking of our AOT scale items outlined in this paper as a caution about how easily ideological biases can infuse our research (Crawford & Jussim, 2018; Duarte et al., 2015). As psychology focuses more and more on charged topics such as the correlates of voting behavior (Pennycook & Rand, 2019), feelings about income inequality (Davidai & Gilovich, 2018; Nero, Swan, Chambers, & Heesacker, 2018), and the psychological correlates of ideology and worldview (Crawford & Jussim, 2018; Crawford & Pilanski, 2014), studies will be more at risk from subtle (even unconscious) contamination by the attitudes of the researchers.

When Richard West and Keith Stanovich began testing out actively open-minded thinking scales in the early 1990s, and when Maggie Toplak joined the team somewhat later, we had no interest at all in political psychology or the study of ideology. We were focused on studying individual differences in performance on heuristics and biases tasks (Stanovich & West, 1998) and on using individual differences to help to adjudicate the great rationality debate in cognitive science (Kahneman & Tversky, 1996; Stanovich & West, 2000; Stanovich, 1999; Tetlock & Mellers, 2002). Yet it appears that our own political/worldview conceptions leaked into these items in subtle ways, nevertheless. The lack of intellectual diversity in our own research team prevented us from seeing how charged the word "belief" might be for a religious person. We provided a specific example of what Reyna (2018), in her discussion of ideological bias in psychology, shows is quite widespread in our field: "Researchers with a liberal bias may be more likely to assume that their interpretation of a survey question means the same thing to all participants and therefore may reach conclusions about the meaning of their findings based on this faulty assumption" (p. 85).

Our original intent when creating the first AOT scales was clearly to generate items with some universality—items equally diagnostic for all potential subjects (at least within the North American context). It never occurred to us that these items would disadvantage any demographic group, let alone the religious minded. No doubt it never occurred to us because not a single member of our lab had any religious inclinations at all.

And, no doubt, the direction of our first observed correlations probably reinforced our biases. From the beginning, AOT scores were negatively correlated with religiosity in our samples. Had the correlations in fact been positive—indicating that the religious-minded were more open-minded—we are sure that we would have pondered much longer and harder about why there was a correlation in the first place. We might have been inclined to think more deeply about each individual item—looking for what was causing the unexpected finding that the religious-minded were more open-minded. But because the correlation was not in that direction, we were not motivated to do that. We were not motivated to look closely at the specific characteristics of items that were more highly correlated with religiosity because the overall correlation was negative, and thus in the direction that we all expected. Had we looked more closely, we would have seen 20 years ago that the belief revision items were disproportionate contributors to the negative correlation that we reliably obtained with religiosity.

In short, what we have done in this paper is what we should have done over 20 years earlier—look more closely at the items in the AOT

scale in order to identify the causes of the negative correlation between religiosity and actively open-minded thinking. We were induced to do so at this late date by some other startlingly high correlations we had observed in the literature. We have empirically demonstrated in this paper the type of item that causes these extra-high correlations and we have confirmed a hypothesis about why these items inflate religiosity correlations.

However, we should be clear that we do consistently get a negative religiosity correlation—it is simply of a much lower magnitude than those in these recent studies. Indeed, when measured correctly, it is of a lower magnitude than what we ourselves observed in earlier versions of our scale. Using the non-BR AOT items, we observed a correlation of −0.25 in study RT60 of the CART (see Table 1). Across the three conditions of the experiment that we reported above (see Table 5), we get correlations with non-BR items of −0.25, −0.09, and −0.22, respectively.

Finally, using items of the Secular BR type that we recommend for future scales, we get a correlation of −0.27 (see Table 5). Thus, there is no doubt that there is such a correlation. It is just that it is in the range of −0.20 to −0.30 rather than −0.65. The difference matters, because of the contexts in which many of the correlations in the range of −0.60 to −0.70 have been obtained. That context has been, in many cases, studies that have used only short forms of the AOT with modest reliabilities. If these −0.65 to −0.70 correlations were corrected for attenuation—or if the two variables were measured as latent constructs—it would not be surprising if the relationship between them approached −1.0.

With individual differences in AOT entirely explained by religiosity, psychological research would then be saying to the public that religiosity and failing to think in an open-minded manner were, for all intents and purposes, the same thing—that being highly religious is virtually synonymous with being closeminded. Our findings, of course, support the weaker conclusion that there is a replicable moderate correlation between actively open-minded thinking and religiosity, but the findings also have important implications for how we should contextualize the finding of a significant correlation. This is because we have demonstrated that the actual content of the focal beliefs in the "beliefs slot" of a belief revision item change the magnitude of the correlation between AOT and religiosity.

We take our case study of the AOT to be another instance in psychology where general psychological relationships were claimed before it was realized that the magnitude of the relationships was not general at all but depended on the idiosyncrasies of the stimuli chosen for examination. In that sense, our findings are similar to the recent rethinking in social psychology about the meaning of a genre of older research on prejudice, intolerance, and warmth of feelings towards various social groups (Crawford, 2018). It was long thought that outgroup prejudice and intolerance were associated with conservative ideology, low intelligence, and low openness to experience. Proponents of the ideological conflict hypothesis (Brandt, Reyna, Chambers, Crawford, & Wetherell, 2014; Chambers, Schlenker, & Collisson, 2013) questioned the generality of these findings by pointing out that the target social groups in these studies (African-Americans, gay individuals, etc.) were often groups that shared ideological affinity with liberals and had conflicting values with conservatives. Thus, the lower outgroup warmth and tolerance shown by conservatives may well have been due to an ideological conflict with the target groups in the earlier studies.

A substantial literature has now accumulated that has put the ideological conflict hypothesis to empirical test. This literature has converged on the conclusion that measures of outgroup tolerance, prejudice, and warmth are more a function of the degree that the values of the target group match or mismatch the values of the subject rather than the psychological characteristics of the subject (Brandt & Crawford, 2019; Chambers et al., 2013; Crawford & Pilanski, 2014; Wetherell, Brandt, & Reyna, 2013). Intolerance and lack of warmth correlations with conservatism, low intelligence, or openness virtually disappear once a more diverse range of social groups are included in the rating task. It seems that liberals too can show relative intolerance—they just express it toward different groups (businesspeople, Christian fundamentalists, the wealthy, the military) that do not share their worldview or values. Our studies here are similar in that we found that correlations with religiosity and ideology were attenuated once AOT stimuli put protected values at risk more equally across the religiosity/ideological spectrum.

It is increasingly the case that the psychological correlates of worldview, voting behavior, and ideological orientation are becoming points of contention in our divided political culture (Baron & Jost, 2019; Baron, 2018; Crawford & Jussim, 2018; Ditto et al., 2019; Duarte et al., 2015; Kahan, 2013; Kahan, Peters, Dawson, & Slovic, 2017; Stanovich, 2017). If psychological studies of this type are becoming an increasing adjunct of politics, it is important that psychology maintain its credibility as a neutral arbiter—a credibility that has been vastly eroded in recent years by empirical evidence of the ideological bias in our science (Ceci & Williams, 2018; Crawford & Jussim, 2018; Duarte et al., 2015). There is a need for greater intellectual diversity in all areas of psychology, but particularly in those that interface with politics and sociocultural beliefs. Greater intellectual diversity in our own lab years ago might have prevented us from continuing to use items in our AOT scale that inflated negative correlations with religiosity. A religiously oriented team member in our lab might have more quickly seen the possibility that the belief revision items made it harder for the religious minded to display actively open-minded thinking when responding. In that sense, our own research on AOT becomes a case study of how ideological bias can infiltrate the work of even the most well-intentioned scientists.

Acknowledgements

This research was supported by a grant from the Social Sciences and Humanities Research Council of Canada to Maggie Toplak (Grant File #: 435-2018-0874).

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://dataverse.scholarsportal.info/dataset.xhtml?persistentId=doi:10.5683/SP2/7R4YQA.

Appendix B. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2019.03.006.

References

Baron, J. (1985). *Rationality and intelligence.* Cambridge: Cambridge University Press.
Baron, J. (1988). *Thinking and deciding* (1st ed.). Cambridge, MA: Cambridge University Press.
Baron, J. (2008). *Thinking and deciding* (4th ed.). Cambridge, MA: Cambridge University Press.
Baron, J. (2015). Supplement to Deppe et al. (2015). *Judgment and Decision Making, 10*(4), 1–2.
Baron, J. (2018). Actively open-minded thinking in politics. *Cognition*.
Baron, J., Gürçay, B., & Metz, S. E. (2017). Reflective thought and actively open-minded thinking. In M. Toplak, & J. Weller (Eds.). *Individual differences in judgment and decision making* (pp. 107–126). London: Routledge.
Baron, J., & Jost, J. T. (2019). False equivalence: Are liberals and conservatives in the U.S. equally "Biased"? *Perspectives on Psychological Science*.
Baron, J., & Leshner, S. (2000). How serious are expressions of protected values? *Journal of Experimental Psychology: Applied, 6*, 183–194.
Baron, J., Scott, S., Fincher, K., & Metz, S. E. (2015). Why does the cognitive reflection test (sometimes) predict utilitarian moral judgment (and other things)? *Journal of Applied Research in Memory and Cognition, 4*, 265–284.
Baron, J., & Spranca, M. (1997). Protected values. *Organizational Behavior and Human Decision Processes, 70*, 1–16.

Bartels, D. M., & Medin, D. L. (2007). Are morally-motivated decision makers insensitive to the consequences of their choices? *Psychological Science, 18*, 24–28.
Brandt, M. J., & Crawford, J. T. (2019). The ideological-conflict hypothesis: Intolerance among both liberals and conservatives. *Current Directions in Psychological Science, 28*.
Brandt, M. J., Reyna, C., Chambers, J. R., Crawford, J. T., & Wetherell, G. (2014). The ideological-conflict hypothesis: Intolerance among both liberals and conservatives. *Current Directions in Psychological Science, 23*, 27–34.
Bronstein, M. V., Pennycook, G., Bear, A., Rand, D. G., & Cannon, T. D. (2018). Belief in fake news is associated with delusionality, dogmatism, religious fundamentalism, and reduced analytic thinking. *Journal of Applied Research in Memory and Cognition*.
Campitelli, G., & Gerrans, P. (2014). Does the cognitive reflection test measure cognitive reflection? A mathematical modeling approach. *Memory & Cognition, 42*, 434–447.
Carl, N. (2014). Verbal intelligence is correlated with socially and economically liberal beliefs. *Intelligence, 44*, 142–148.
Ceci, S. J., & Williams, W. M. (2018). Socio-political values infiltrate the assessment of scientific research. In J. Crawford, & L. Jussim (Eds.). *The politics of social psychology* (pp. 156–167). New York: Routledge.
Cederblom, J. (1989). Willingness to reason and the identification of the self. In E. Maimon, B. Nodine, & F. O'Conner (Eds.). *Thinking, reasoning, and writing* (pp. 147–159). New York: Longman.
Chambers, J. R., Schlenker, B. R., & Collisson, B. (2013). Ideology and prejudice: The role of value conflicts. *Psychological Science, 24*, 140–149.
Crawford, J. T. (2018). The politics of the psychology of prejudice. In J. Crawford, & L. Jussim (Eds.). *The politics of social psychology* (pp. 99–115). New York: Routledge.
Crawford, J. T., & Jussim, L. (2018). *The politics of social psychology*. New York: Routledge.
Crawford, J. T., & Pilanski, J. (2014). Political intolerance, right *and* left. *Political Psychology, 35*, 841–851.
Davidai, S., & Gilovich, T. (2018). How should we think about Americans' beliefs about economic mobility? *Judgment and Decision Making, 13*, 297.
De Neys, W. (2014). Conflict detection, dual processes, and logical intuitions: Some clarifications. *Thinking & Reasoning, 20*, 169–187.
Deniz, H., Donnelly, L. A., & Yilmaz, I. (2008). Exploring the factors related to acceptance of evolutionary theory among Turkish preservice biology teachers. *Journal of Research in Science Teaching, 45*, 420–443.
Dienes, Z., & Perner, J. (1999). A theory of implicit and explicit knowledge. *Behavioral and Brain Sciences, 22*, 735–808.
Ditto, P. H., Liu, B., Clark, C., Wojcik, S., Chen, E., Grady, R., et al. (2019). At least bias is bipartisan: A meta-analytic comparison of partisan bias in liberals and conservatives. *Perspectives on Psychological Science*.
Duarte, J. L., Crawford, J. T., Stern, C., Haidt, J., Jussim, L., & Tetlock, P. E. (2015). Political diversity will improve social psychological science. e130 *Behavioral and Brain Sciences, 38*(1–58), https://doi.org/10.1017/S0140525X14000430.
Evans, J. St. B. T. (2002). The influence of prior belief on scientific thinking. In P. Carruthers, S. Stich, & M. Siegal (Eds.). *The cognitive basis of science* (pp. 193–210). Cambridge: Cambridge University Press.
Evans, J. St. B. T. (2007). *Hypothetical thinking: Dual processes in reasoning and judgment.* New York: Psychology Press.
Evans, J. St. B. T. (2010). *Thinking twice: Two minds in one brain.* Oxford: Oxford University Press.
Evans, J. St. B. T. (2017). Belief bias in deductive reasoning. In R. Pohl (Ed.). *Cognitive illusions* (pp. 165–181). (2nd ed.). London: Routledge.
Evans, J. St. B. T., & Over, D. E. (1999). Explicit representations in hypothetical thinking. *Behavioral and Brain Sciences, 22*, 763–764.
Evans, J. St. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science, 8*, 223–241.
Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives, 19*, 25–42.
Gervais, W. M., & Norenzayan, A. (2012). Analytic thinking promotes religious disbelief. *Science, 336*(6080), 493–496.
Gervais, W. M., van Elk, M., Xygalatas, D., McKay, R. T., Aveyard, M., Buchtel, E., et al. (2018). Analytic atheism: A cross-culturally weak and fickle phenomenon? *Judgment and Decision Making, 13*, 268–274.
Goldman, A. I. (1986). *Epistemology and cognition.* Cambridge, MA: Harvard University Press.
Hahn, U., & Harris, A. J. L. (2014). What does it mean to be biased: Motivated reasoning and rationality. *The Psychology of Learning and Motivation, 61*, 41–102.
Haran, U., Ritov, I., & Mellers, B. A. (2013). The role of actively open-minded thinking in information acquisition, accuracy, and calibration. *Judgment and Decision Making, 8*, 188–201.
Harman, G. (1995). Rationality. In E. E. Smith, & D. N. Osherson (Eds.). *Thinking* (pp. 175–211). Cambridge, MA: The MIT Press.
Heijltjes, A., van Gog, T., Leppink, J., & Paas, F. (2014). Improving critical thinking: Effects of dispositions and instructions on economics students' reasoning skills. *Learning and Instruction, 29*, 31–42.
Heijltjes, A., van Gog, T., Leppink, J., & Paas, F. (2015). Unraveling the effects of critical thinking instructions, practice, and self-explanation on students' reasoning performance. *Instructional Science, 43*, 487–506.
Kahan, D. (2013). Ideology, motivated reasoning, and cognitive reflection. *Judgment and Decision Making, 8*, 407–424.
Kahan, D., Peters, E., Dawson, E., & Slovic, P. (2017). Motivated numeracy and enlightened self-government. *Behavioural Public Policy, 1*, 54–86.
Kahneman, D. (2011). *Thinking, fast and slow.* New York: Farrar, Straus & Giroux.
Kahneman, D., & Tversky, A. (1996). On the reality of cognitive illusions. *Psychological Review, 103*, 582–591.
Kelman, M. (2011). *The heuristics debate.* New York: Oxford University Press.
Klayman, J., & Ha, Y. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review, 94*, 211–228.
Kokis, J., Macpherson, R., Toplak, M., West, R. F., & Stanovich, K. E. (2002). Heuristic and analytic processing: Age trends and associations with cognitive ability and cognitive styles. *Journal of Experimental Child Psychology, 83*, 26–52.
Kovacs, K., & Conway, A. (2016). Process overlap theory: A unified account of the general factor of intelligence. *Psychological Inquiry, 27*, 151–177.
McKenzie, C. R. M. (2004). Hypothesis testing and evaluation. In D. J. Koehler, & N. Harvey (Eds.). *Blackwell handbook of judgment & decision making* (pp. 200–219). Malden, MA: Blackwell.
Mellers, B., Stone, E., Atanasov, P., Rohrbaugh, N., Metz, S. E., Ungar, L., et al. (2015). The psychology of intelligence analysis: Drivers of prediction accuracy in world politics. *Journal of Experimental Psychology: Applied, 21*, 1–14.
Mercier, H. (2017). Confirmation bias – myside bias. In R. Pohl (Ed.). *Cognitive illusions* (pp. 99–114). (2nd ed.). London: Routledge.
Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions four general conclusions. *Current Directions in Psychological Science, 21*, 8–14.
Nero, S. S., Swan, L. K., Chambers, J. R., & Heesacker, M. (2018). Still no compelling evidence that Americans overestimate upward socio-economic mobility rates: Reply to Davidai & Gilovich (2018). *Judgment and Decision Making, 13*, 305–308.
Nigg, J. T. (2017). On the relations among self-regulation, self-control, executive functioning, effortful control, cognitive control, impulsivity, risk-taking, and inhibition for developmental psychopathology. *Journal of Child Psychology and Psychiatry, 58*, 361–383.
Nozick, R. (1993). *The nature of rationality.* Princeton, NJ: Princeton University Press.
Oaksford, M., & Chater, N. (2012). Dual processes, probabilities, and cognitive architecture. *Mind & Society, 11*, 15–26.
Palan, S., & Schitter, C. (2018). Prolific.ac—a subject pool for online experiments. *Journal of Behavioral and Experimental Finance, 17*, 22–27.
Paolacci, G., & Chandler, J. (2014). Inside the Turk: Understanding Mechanical Turk as a participant pool. *Current Directions in Psychological Science, 23*, 184–188.
Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology, 70*, 153–163.
Pennycook, G., Cheyne, J. A., Barr, N., Koehler, D. J., & Fugelsang, J. A. (2014). Cognitive style and religiosity: The role of conflict detection. *Memory & Cognition, 42*, 1–10.
Pennycook, G., & Rand, D. G. (2019). Cognitive reflection and the 2016 US presidential election. *Personality and Social Psychology Bulletin, 45*, 224–239.
Piazza, J., & Landy, J. (2013). "Lean not on your own understanding": Belief that morality is founded on divine authority and non-utilitarian moral thinking. *Judgment and Decision Making, 8*, 639–661.
Pronin, E. (2007). Perception and misperception of bias in human judgment. *Trends in Cognitive Sciences, 11*, 37–43.
Reyna, C. (2018). Scale creation, use, and misuse: How politics undermines measurement. In J. Crawford, & L. Jussim (Eds.). *The politics of social psychology* (pp. 81–98). New York: Routledge.
Sá, W., Kelley, C., Ho, C., & Stanovich, K. E. (2005). Thinking about personal theories: Individual differences in the coordination of theory and evidence. *Personality and Individual Differences, 38*, 1149–1161.
Sá, W., & Stanovich, K. E. (2001). The domain specificity and generality of mental contamination: Accuracy and projection in judgments of mental content. *British Journal of Psychology, 92*, 281–302.
Sá, W., West, R. F., & Stanovich, K. E. (1999). The domain specificity and generality of belief bias: Searching for a generalizable critical thinking skill. *Journal of Educational Psychology, 91*, 497–510.
Samuelson, P. L., & Church, I. M. (2015). When cognition turns vicious: Heuristics and biases in light of virtue epistemology. *Philosophical Psychology, 28*, 1095–1113.
Sinatra, G. M., Southerland, S. A., McConaughy, F., & Demastes, J. W. (2003). Intentions and beliefs in students' understanding and acceptance of biological evolution. *Journal of Research in Science Teaching, 40*, 510–528.
Stanovich, K. E. (1999). *Who is rational? Studies of individual differences in reasoning.* Mahwah, NJ: Erlbaum.
Stanovich, K. E. (2004). *The robot's rebellion: Finding meaning in the age of Darwin.* Chicago: University of Chicago Press.
Stanovich, K. E. (2008). Higher-order preferences and the master rationality motive. *Thinking & Reasoning, 14*, 111–127.
Stanovich, K. E. (2011). *Rationality and the reflective mind.* New York: Oxford University Press.
Stanovich, K. E. (2012). On the distinction between rationality and intelligence: Implications for understanding individual differences in reasoning. In K. Holyoak, & R. Morrison (Eds.). *The Oxford handbook of thinking and reasoning* (pp. 343–365). New York: Oxford University Press.
Stanovich, K. E. (2017, September 28). Were Trump voters irrational? Quillette. Retrieved from http://quillette.com/2017/09/28/trump-voters-irrational/.
Stanovich, K. E. (2018). Miserliness in human cognition: The interaction of detection, override, and mindware. *Thinking & Reasoning*, 24.
Stanovich, K. E., & Toplak, M. E. (2012). Defining features versus incidental correlates of Type 1 and Type 2 processing. *Mind & Society, 11*, 3–13.
Stanovich, K. E., & West, R. F. (1997). Reasoning independently of prior belief and individual differences in actively open-minded thinking. *Journal of Educational Psychology, 89*, 342–357.
Stanovich, K. E., & West, R. F. (1998). Individual differences in rational thought. *Journal of Experimental Psychology: General, 127*, 161–188.
Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences, 23*, 645–726.
Stanovich, K. E., & West, R. F. (2007). Natural myside bias is independent of cognitive

ability. *Thinking & Reasoning, 13*, 225–247.
Stanovich, K. E., & West, R. F. (2008). On the failure of intelligence to predict myside bias and one-sided bias. *Thinking & Reasoning, 14*, 129–167.
Stanovich, K. E., West, R. F., & Toplak, M. E. (2013). Myside bias, rational thinking, and intelligence. *Current Directions in Psychological Science, 22*, 259–264.
Stanovich, K. E., West, R. F., & Toplak, M. E. (2016). *The rationality quotient: Toward a test of rational thinking.* Cambridge, MA: MIT Press.
Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin, 87*, 245–251.
Stenhouse, N., Myers, T., Vraga, E., Kotcher, J., Beall, L., & Maibach, E. (2018). The potential role of actively open-minded thinking in preventing motivated reasoning about controversial science. *Journal of Environmental Psychology, 57*, 17–24.
Svedholm, A. M., & Lindeman, M. (2013). The separate roles of the reflective mind and involuntary inhibitory control in gatekeeping paranormal beliefs and the underlying intuitive confusions. *British Journal of Psychology, 104*, 303–319.
Svedholm-Häkkinen, A. M., & Lindeman, M. (2018). Actively open-minded thinking: Development of a shortened scale and disentangling attitudes towards knowledge and people. *Thinking & Reasoning, 24*, 21–40.
Swami, V., Voracek, M., Stieger, S., Tran, U. S., & Furnham, A. (2014). Analytic thinking reduces belief in conspiracy theories. *Cognition, 133*, 572–585.
Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science, 50*, 755–769.
Tetlock, P. E., & Mellers, B. A. (2002). The great rationality debate. *Psychological Science, 13*, 94–99.
Toplak, M. E., West, R. F., & Stanovich, K. E. (2011). The cognitive reflection test as a predictor of performance on heuristics and biases tasks. *Memory & Cognition, 39*, 1275–1289.
Toplak, M. E., West, R. F., & Stanovich, K. E. (2014). Rational thinking and cognitive sophistication: Development, cognitive abilities, and thinking dispositions. *Developmental Psychology, 50*, 1037–1048.
Wetherell, G. A., Brandt, M. J., & Reyna, C. (2013). Discrimination across the ideological divide: The role of value violations and abstract values in discrimination by liberals and conservatives. *Social Psychological and Personality Science, 4*, 658–667.
West, R. F., Toplak, M. E., & Stanovich, K. E. (2008). Heuristics and biases as measures of critical thinking: Associations with cognitive ability and thinking dispositions. *Journal of Educational Psychology, 100*, 930–941.
Yilmaz, O., & Saribay, S. A. (2016). An attempt to clarify the link between cognitive style and political ideology: A non-Western replication and extension. *Judgment and Decision Making, 11*, 287–300.
Yilmaz, O., & Saribay, S. A. (2017). The relationship between cognitive style and political orientation depends on the measures used. *Judgment and Decision Making, 12*, 140–147.