

ORIGINAL ARTICLE

Analytic atheism and analytic apostasy across cultures

Nick Byrd¹ , Stephen Stich²  and Justin Sytsma³ 

¹Department of Bioethics and Decision Sciences, Geisinger College of Health Sciences, Danville, Pennsylvania, USA; ²Department of Philosophy, Rutgers University, New Brunswick, New Jersey, USA and ³Department of Philosophy, Victoria University of Wellington, Wellington, New Zealand
Corresponding author: Nick Byrd; Email: nbyrd@geisinger.edu

(Received 30 October 2024; revised 7 March 2025; accepted 10 March 2025; first published online 2 April 2025)

Abstract

Reflective thinking often predicts less belief in God or less religiosity – so-called analytic atheism. However, those correlations involve limitations: widely used tests of reflection confound reflection with ancillary abilities such as numeracy; some studies do not detect analytic atheism in every country; experimentally encouraging reflection makes some non-believers more open to believing in God; and one of the most common online research participant pools seems to produce lower data quality. So analytic atheism may be less than universal or partially explained by confounding factors. To test this, we developed better measures, controlled for more confounds, and employed more recruitment methods. All four studies detected signs of analytic atheism above and beyond confounds ($N > 70,000$ people from five of six continental regions). We also discovered analytic apostasy: the better a person performed on reflection tests, the greater their odds of losing their religion since childhood – even when controlling for confounds. Analytic apostasy even seemed to explain analytic atheism: apostates were more reflective than others and analytic atheism was undetected after excluding apostates. Religious conversion was rare and unrelated to reflection, suggesting reflection's relationships to conversion and deconversion are asymmetric. Detected relationships were usually small, indicating reflective thinking is a reliable albeit marginal predictor of apostasy.

Keywords: cognitive science of religion; experimental philosophy; dual process theory; analytic atheism; apostasy

Introduction

Reflective reasoning is the familiar phenomenon of backing up from an initial impulse to reappraise it in light of reasons and alternatives (Korsgaard 1996). Philosophers tend to think that reflective reasoning is good: reflection is supposed to correct faulty impulses or find reasons that justify beliefs that we had not yet questioned (Sosa 2009; cf. Byrd 2022). Sure enough, correct answers on reflection tests predict better reasoning about logic (Byrd and Conway 2019), probability (Liberali et al. 2012), and physics (Gette and Kryjevskaja, 2019). Not surprisingly, when some scientists found that unreflective thinking correlated with theism and reflective thinking correlated with atheism (Shenhav et al. 2012) and other scientists labelled these results with the name ‘analytic atheism’ (Norenzayan and Gervais 2012, Table 2), philosophers quickly noted how the name ‘earnestly congratulates

religion on its lack of cognitive content' (Pigden 2013: 312). Of course, anyone who has studied religion knows that highly reflective believers exist. As such they may wonder whether these reflective religionists simply do not make it into the studies that find reflective thinking correlating with atheism or agnosticism (Pennycook *et al.* 2016). However, even studies including academic philosophers find moderate correlations between reflection and atheism (Byrd 2023). This suggests that links between reflection and areligiosity may be somewhat prevalent.

However, links between reflection and areligiosity are not ubiquitous. Despite finding the correlation between reflection and disbelief *across* about a dozen countries ($N = 3461$), researchers do not find it *within* each of those countries (Gervais *et al.* 2018). And when studying causal relationships between reflection and religiosity, the so-called analytic atheist effect is not always observed (Saribay *et al.* 2020; Yilmaz and Isler 2019). Worse, there are enough problems with standard protocols for measuring and manipulating reflection that we may need to reconsider whether many prior results actually support the conclusion that atheism is linked to analytic or reflective thinking (Byrd *et al.* 2023).

To address these mixed results and methodological concerns, we developed better measures of religiosity, reflection, and known confounds. In multiple large studies, people from around the world exhibited signs of analytic atheism and even analytic *apostasy*. Data were filtered and analysed in Jamovi to allow readers to reproduce our analyses without any coding experience or paywalled software. All collected data and exclusions are reported. Pre-registration, data, analysis files, and appendix are available on the Open Science Framework: <https://osf.io/8wf43/>

Study I

Prior to pre-registering new research designs, hypotheses, and analyses, we wanted to test whether an analytic atheist correlation could be found in a large, culturally diverse sample while controlling for potentially confounding factors. The idea was that if the expected relationship were found in this high-powered dataset, it would be worth pre-registering more sophisticated studies.

Method

From 2009 to 2018, a Google ad invited people to take surveys in return for personality test results – also known as the ‘push out’ method, which has been shown to yield more demographically diverse samples than ‘pull in’ methods such as Amazon Mechanical Turk, CloudResearch, or Prolific (Antoun *et al.* 2016: 232). Data were collected for a large collection of other studies (Feltz and Cokely 2011; Livengood *et al.* 2010; Machery *et al.* 2017; Murray *et al.* 2013), but this article is the first to fully aggregate and analyse the data for insights about analytic atheism.

Participants

Of the people who clicked the advertisement and consented to participate, 71,591 completed versions of the survey that included measures of our target variables (religiosity and reflection). To mitigate the impact of low data quality or low power, we excluded from analysis participants who reported at least one of the following: an age of 100 or more ($n = 28$), a wildly implausible answer to any reflection test question ($n = 3124$), a country that is either insincere (such as ‘Earth’ or ‘Agrabah’) or that fewer than 100 other participants reported ($n = 2819$). The following analyses are based on the remaining sample of 65,873 responses.

Materials

Participants took a reflection test and answered questions about religiosity, personality, politics, as well as demographics. Descriptive statistics for Study 1 are in [Table 1](#).

Table 1. Descriptive statistics for study 1

	Category	<i>n</i>	Per cent
Gender	Female	46632	71%
	Male	19241	29%
Race	Asian or Pacific Islander	3091	5%
	African American	8739	13%
	Hispanic	6436	10%
	Native American or Alaskan	1613	2%
	Caucasian	45994	70%
Country	Canada	366	1%
	Philippines	196	< 1%
	India	560	1%
	Australia	243	< 1%
	South Africa	176	< 1%
	Nigeria	117	< 1%
	United States	63663	97%
	United Kingdom	508	1%
First language	English	60635	92%
	Not English	5238	8%
	None	11484	17%
	Catholic	10613	16%
	Protestant	9189	14%
	Mormon	1081	2%
Religion	Other Christian	19858	30%
	Jewish	720	1%
	Muslim	530	1%
	Hindu	609	1%
	Buddhist	562	1%
	Pagan	582	1%

(Continued)

Table 1. (Continued.)

	Category			n	Per cent	
	Atheist			1657	3%	
	Agnostic			2515	4%	
	Other			6463	10%	
	Mean	Median	Mode	SD	Minimum	Maximum
Age	36.8	34	18	15.1	18	100
Income	2.8	3	1	1.6	1	8
Political Orientation	4.0	4	4	1.7	1	7
Educational Attainment	4.0	4	4	1.8	1	7
University Training in Philosophy	0.2	0	0	0.6	0	5
University training in Psychology	0.4	0	0	0.8	0	5
Religiosity	2.7	3	3	1.4	1	5
Openness	10.9	11	12	2.5	2	14
Conscientiousness	10.8	11	14	2.8	2	14
Extraversion	8.7	9	8	3.3	2	14
Agreeableness	10.1	10	9	2.6	2	14
Neuroticism	9.1	9	8	3.2	2	14
Bat and ball answer	10.8	10	10	154.8	0	10500
Widgets answer	90.2	100	100	107.0	0	500
Lily pad answer	27.5	24	24	10.8	0	48

Demographics. Participants were asked to report their age, gender, income (1 = ‘Less than \$10,000,’ 3 = ‘\$25,000 to \$50,000,’ 8 = ‘More than \$250,000’), political orientation (1 = ‘Very liberal,’ 4 = ‘Neither liberal nor conservative,’ 7 = ‘Very conservative’), and country.

Education. Participants selected their educational attainment (1 = ‘Some High School,’ 4 = ‘Some College,’ 7 = ‘Graduate or Professional Degree’), as well as university education in Philosophy (0 = ‘Some Undergraduate Courses’ to 5 = ‘PhD’) and Psychology (on the same scale).

First language. Our survey was written in English, but Google ads may appear to users who are not fluent users of English who may choose answers without fully understanding

the question or their answer. To control for this measurement error, participants were asked to report if English was their first language.

Religiosity. Participants answered, 'If religiosity is defined as participating with an organized religion, then to what degree do you consider yourself religious?' on a scale from 'Not at all' (1) to 'Totally' (5), with 'Somewhat' at the midpoint (3). They also reported their religion.

Ten-Item Personality Inventory (TIPI). Participants also rated their agreement with 10 descriptions of their personality such as 'Dependable, Self-disciplined' or 'Extraverted, Enthusiastic' on a scale from 'Disagree Strongly' (1) to 'Agree Strongly' (7). Pairs of scores were summed to produce scores for five traits: Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism.

Cognitive Reflection Test. Participants completed the original, three-item Cognitive Reflection Test (Frederick 2005). Each item is designed to lure participants to choose an answer that seems correct but that – upon reflection – is demonstrably incorrect (Byrd 2019). Consider an example. 'A bat and a ball cost \$1.10 in total. The bat costs \$1.00 more than the ball. How much does the ball cost? ___ cents.' The lured answer is 10, but the correct answer is 5. Correct answers on the test were summed ($\alpha = 0.63$). So were lured responses ($\alpha = 0.38$), which were the most common responses for each item.

Results

Linear regression was used to predict religiosity. Figure 1 shows a small negative correlation between religiosity and correct reflection test answers ($\beta = -0.08$, 95% CI $[-0.09, -0.07]$, $p < 0.001$) in the top left, controlling for lured reflection test answers ($\beta = 0.00$, 95% CI $[-0.01, 0.01]$, $p = 0.641$). After controlling for other measured variables, reflection's relationship with religiosity remained ($\beta = -0.06$, 95% CI $[-0.08, -0.05]$, $p < 0.001$), was negative in most countries (top right), and independent of education (bottom left) and even training in philosophy (bottom right).

The Appendix's Table A1 reports the results of the full model, which accounts for 13% of the variance in religiosity – adjusted $R^2 = 0.13$, $F(34, 65794) = 286.49$, $p < 0.001$. Collinearity statistics are in Table A2.

Discussion

Study 1 conceptually replicated analytic atheist correlations in a large culturally diverse sample. Reflection test performance predicted a small amount of variance (1%) in religiosity across eight countries and country-by-country relationships between religiosity and reflection were usually negative.

Of course, there are limitations to what we can infer from Study 1. Crucially, we cannot necessarily infer from these correlations *between* people that *within* each person (on average) religiosity decreases as reflection increases (Fisher et al. 2018). And since Study 1 was run, limitations about some of the measures have emerged. Mounting evidence has suggested this original three-item reflection test confounds reflection with math ability (Attali and Bar-Hillel 2020; Erceg et al. 2020). Worse, studies have found that correlations between philosophical inclinations and this mathematical reflection test may have more to do with math ability than reflective thinking (Byrd and Conway 2019). Moreover, the US-centric or Western-centric terminology in this survey may mask nuance. For example, one-dimensional questions about political orientation overlook how variables such as religiosity and reflection can relate to *social* conservatism differently than they do to *economic* conservatism (Saribay and Yilmaz 2018; Yilmaz et al. 2020). So we sought funding for follow-up studies that employ more suitable study design and measures.

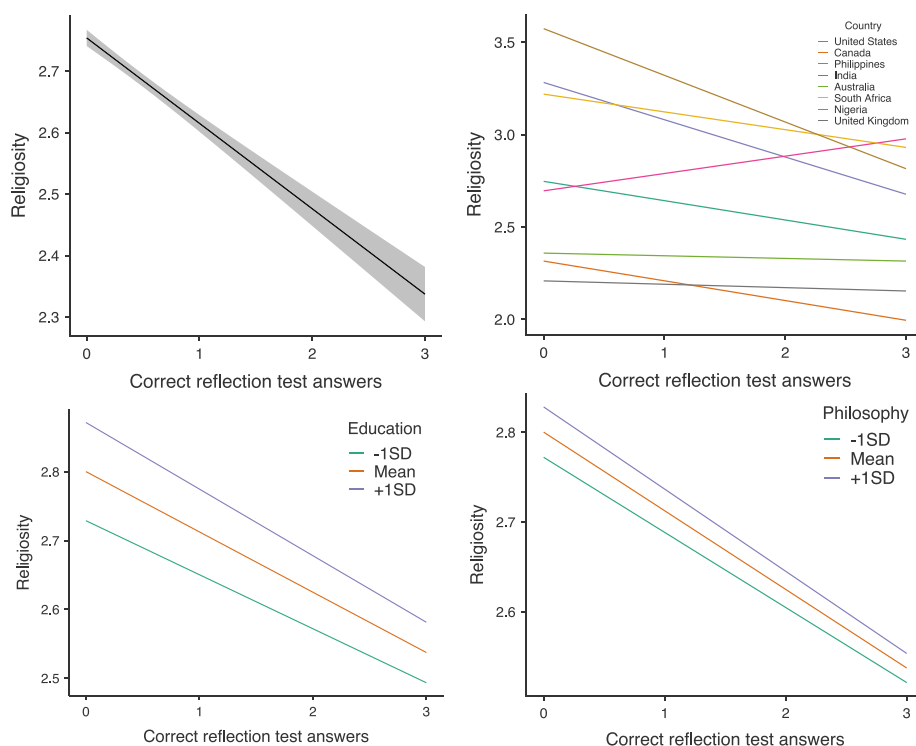


Figure 1. Religiosity by correct test answers with 95% confidence intervals (top left), controlling for lured answers and then controlling for all variables by country (top right), education (bottom left) and philosophy (bottom right).

Study 2

Study 2 aimed to develop better materials to test reflection (not just math ability), detect changes in religiosity, and measure potentially confounding factors.

Method

In Study 2, we employed a ‘pull’ recruitment method (Antoun *et al.* 2016) to validate better measures of reflection, religiosity, demographics, and education with participants from the United States.

Participants

We pulled in only those Amazon Mechanical Turk (MTurk) workers that passed a battery of CloudResearch’s quality controls (Litman *et al.* 2023) since MTurk’s internal metrics (such as approval rate or minimum submissions) are not good indicators of quality (Byrd 2025; Hauser *et al.* 2023). We aimed to recruit 250 people to allow observational relationships to stabilize (Schönbrodt and Perugini 2013). The following analyses exclude only respondents who did not complete the required survey questions ($n = 8$) or failed an attention check ($n = 16$), leaving a final sample of 251. Compensation was \$3.00 USD.

Materials

Participants took a novel reflection test and answered questions about religion, politics, education as well as demographics. Table 2 shows descriptive statistics for demographics, education, reflection, and religion for Study 2.

Demographics. After participants reported birthyear and gender,¹ they indicated political orientation, with some additional response options (compared to Study 1). As predicted by research conducted since data for Study 1 were collected (e.g., Yilmaz et al. 2020), 37 of our participants from Study 2 (15 per cent) selected ‘Don’t know’, ‘Libertarian’, or ‘Other’ (–1). To overcome potential limitations of the one-dimensional political scale, participants were also able to indicate both Social and Economic conservatism on the same scale from ‘Very liberal’ (1) to ‘Very conservative’ (without additional options). Participants also reported the country in which they were raised, the religion in which they were raised, and their current religion. Due to underrepresentation of non-Christian religions in this sample, they were collapsed into an ‘Other Religion’ bucket to maximize their statistical power in our analyses.

General education. Participants selected their ‘highest level of education’ ranging from ‘Less than high school degree’ (1) to ‘Doctoral ...’ or ‘Professional degree (J.D., M.D.)’ (7) with an option to report ‘Other’ (0), which eight (3 per cent) participants selected, usually because they reported attending technical, trade, or vocational school. ‘Some university but no degree’ (3) was the most common response. We also asked for parents’ highest levels of education using the same scale, yielding a median response of ‘Associates degree (or 2-year degree)’ (4) and a modal response of ‘High school graduate (or high school diploma equivalent)’ (2). To pilot a more cross-culturally robust measure of education we asked for the number of years that participants had been ‘a student (including prior to the university level, at the university-level, and above the university-level)’. We also asked for the number of years one’s parents were students. As expected, number of years as a student strongly predicted educational attainment – $r > 0.38$, $p < 0.001$ (Figure A3).

Domain-specific education. To control for confounds with reflection test performance, we also asked participants to report the number of STEM and philosophy courses they have taken. Likewise, we asked, ‘Have you ever studied critical thinking?’ with ‘Yes’ (1) and ‘No’ (0) response options.

Abbreviated religiosity scale. We piloted a short but broad religiosity scale to maximize response quality in subsequent studies by using fewer items than many validated religiosity scales contain (Galesic and Bosnjak 2009). The abbreviated scale exhibited excellent reliability ($\alpha = 0.89$). Moreover, all but the Superstition item loaded on one factor (loadings > 0.4 , Bartlett’s $\chi^2(66) = 1710.83$, $p < 0.001$).

Religious belief and identity. Participants rated their agreement with six items on a scale from ‘Strongly disagree’ (–2) to ‘Strongly agree’ (2) with ‘Neither agree nor disagree’ at the midpoint (0).

Religious practice. Participants also reported the frequency of attending religious events and practicing religious disciplines on a scale from ‘Never’ (0) to ‘More than once per week’ (5).

Religion’s importance. Participants rated the importance of religious or spiritual community and belief on a scale from ‘Highest importance’ (5) to ‘Irrelevant’ (0), with a ‘Not applicable’ option (–1), which 48 participants (19 per cent) selected.

Religion’s influence on life and morality. Given how much some religions moralize their religious norms (Levine et al. 2021), participants rated how much their religious or spiritual beliefs influenced their ‘life decisions’ and ‘moral beliefs’ on a scale from ‘A great deal’ (2) to ‘Not at all’ (0), with 41 participants (16 per cent) selecting ‘I don’t have religious or spiritual beliefs.’

Table 2. Descriptive statistics for study 2

Category				n	Per cent	
Gender	Woman			160	64%	
	Man			84	33%	
	Other			7	3%	
Country	United States			251	100%	
	Christian			84	33%	
Religion	No Religion			114	45%	
	Other Religion			43	17%	
	Don't Know			10	4%	
	Mean	Median	Mode	SD	Minimum	Maximum
Survey Progress	99.18	100	100	0.99	98	100
Survey Duration (s)	962.61	716	570	817.22	216	7141
Survey Experience	1.10	1	1	0.66	0	2
Birthyear	1987.80	1990	2001	10.67	1949	2003
Political orientation	2.77	3	4	2.22	−1	7
Social conservatism	3.29	3	4	1.72	1	7
Econ. conservatism	3.71	4	4	1.67	1	7
Highest Edu. (self)	3.72	3	3	1.5	0	7
Highest Edu. (parent)	3.67	4	2	1.76	0	7
School years (self)	13.43	15	16	6.3	0	30
School years (parent)	12.41	13	12	6.51	0	38
Philosophy courses	1.14	1	0	1.53	0	12
STEM courses	7.02	4	0	8.79	0	50
Crit.Thinking courses	0.46	0	0	0.5	0	1
Nurse item (correct)	0.27	0	0	0.45	0	1
Tea item (correct)	0.42	0	0	0.49	0	1
Race item (correct)	0.45	0	0	0.5	0	1
			Mean	Median	SD	Range
Religious person: 'I am a religious person.'			−0.23	0	1.46	−2 − 2
Spiritual person: 'I am a spiritual person.'			0.58	1	1.34	−2 − 2
Superstitious person: 'I am a superstitious person.'			−0.17	0	1.28	−2 − 2

(Continued)

Table 2. (Continued.)

	Mean	Median	SD	Range
Religious belief: 'I believe in things such as a God or gods, an afterlife, reincarnation, or similar ideas.'	0.75	1	1.36	-2 - 2
Religious reflection: 'It is important to think critically about my religious or spiritual beliefs or practices.'	0.71	1	1.21	-2 - 2
Religious exclusivism: 'The only acceptable religion is my religion.'	-1.37	-2	1.13	-2 - 2
Religious attendance: 'How often do you participate in religious services, celebrations, meetings, or similar religious events?'	1.57	1	1.44	0 - 5
Religious disciplines: 'Outside of participating in religious services, celebrations, meetings, or the like, how often do you engage in religious practices, such as praying, meditating, fasting, volunteering, donating money, or similar activities?'	2.75	3	1.88	0 - 5
Religious community import: How important are 'the religious or spiritual communities with which you are affiliated?'	1.73	2	1.8	-1 - 5
Religious belief import: How important are 'your religious or spiritual beliefs?'	2.62	3	1.83	-1 - 5
Religion influences life: 'How much do your religious or spiritual beliefs influence your life decisions?'	0.74	1	1.01	-1 - 5
Religion influences morality: 'How much do your religious or spiritual beliefs influence your moral beliefs?'	0.93	1	1.09	-1 - 5

Belief in the supernatural. Participants also selected up to 26 beliefs in supernatural entities, processes, or powers. Although unanalysed, the data are openly available.

Long religiosity scales. To gauge how our abbreviated religion scale items correlated with some validated religion scales, we had participants complete some validated religion scales. Reliability was high for all scales. As such, each scales' scores were averaged (Table A3).

Our abbreviated religion items correlated with three extended religion scales that track broad religious belief and commitment (Figure A2 in Appendix). As intended, all abbreviated items except 'Superstitious person' correlated with both the 12-item Religious Worldview scale (Goplen and Plant 2015, Appendix) and the modified six-item intrinsic spirituality scale (Hodge 2003). Some of our abbreviated items correlated with the seven-item Intrinsic Religiosity scale (Tiliopoulos et al. 2006). As expected, our abbreviated items did not correlate with narrow religious constructs like those measured by the 12-item Religious Fundamentalism (Altemeyer and Hunsberger 2004) or seven-item Extrinsic Religiosity scales (Tiliopoulos et al. 2006). Also expected, the extended religion scales did

not correlate as well with one another as our abbreviated religious items did. Indeed, only Intrinsic and Extrinsic religiosity ($r = 0.85$, $p < 0.001$) seemed related. Less expected was that our 'Religious reflection' item did not correlate with the Quest Religiosity scale (Batson and Schoenrade 1991). Overall, these data confirmed that our abbreviated religious items measured what we intended using fewer items than previously validated scales. Given these findings further analysis focused on the abbreviated rather than extended scales.

Novel reflection test. To dissociate reflection from reflection test familiarity (Byrd 2023), we used novel adaptations of the 'nurse', 'race', and 'tea' test questions from Calvillo et al. (2023, Appendix and Supplement). As intended, participants' perceived test familiarity was not related to their performance. To dissociate reflection from numeracy, one of these items was less mathematical: 'You are participating in a race. You pass the person in 3rd place. What place are you in now?' We employed a validated four-option response format (Sirota and Juanchich 2018) including the correct answer (e.g., 3rd), the lure (e.g., 2nd), and two incorrect answers (e.g., 1st or 4th). We summed correct responses ($\alpha = 0.6$) and lured responses ($\alpha = 0.6$), each of which loaded onto a single factor (loadings > 0.6 , Bartlett's $\chi^2(3) = 93.45$, $p < 0.001$), ignoring other incorrect answers.²

Attention check. To further mitigate the impact of low data quality on results, we embedded an instructional attention check into our religion scale: 'Select "strongly agree" for this item' (Kung et al. 2018). Before starting the survey, participants were told it contained this kind of check.

Survey experience. The final required question was, 'Overall, how positive was your experience of this survey?' with response options ranging from 'Extremely negative' (-2) to 'Extremely positive' with 'Neither negative nor positive' at the midpoint (0).

Results

In addition to analytic atheism Study 2 revealed a potentially novel result: analytic apostasy.

Analytic Atheism. Like Study 1, we found small bivariate correlations between answers to our religiosity items and performance on our novel reflection test. Correct reflection test responses predicted less identification as a religious or spiritual person, belief in religious phenomena (e.g., God, an afterlife, reincarnation, etc.), practising of religious or spiritual disciplines, valuing of religious community or belief, and influence of religion or spirituality on life or moral decisions (Figure 2, top). Also, there were small correlations between overconfidence (i.e., perceived rate of correct test answers minus the actual rate of correct answers) and both identifying as religious and practicing religious disciplines, $r \geq 0.13$, $p < 0.014$. No other religious, spiritual, or superstitious items correlated with overconfidence ($r < |0.10|$, $p > 0.139$).

Analytic Apostasy. Binomial logistic regression was used to predict apostasy. Correct reflection test answers predicted 1.6 times higher odds of losing one's childhood religion (OR = 1.55, 95% CI [1.21, 2.00], $p < 0.001$). The bottom left plot of Figure 2 illustrates how reflection's relationship with religiosity remained even after controlling for other measured variables (OR = 1.48, 95% CI [1.10, 1.99], $p = 0.009$). The bottom right of Figure 2 shows how replacing correct with lured reflection test responses predicted the opposite³: as individuals' lured answers increased the odds of losing one religion decreased by more than 30 per cent (OR = 0.65, 95% CI [0.50, 0.83], $p < 0.001$), even when controlling for other factors (OR = 0.68, 95% CI [0.51, 0.92], $p = 0.012$). Reflection also predicted at least as extreme odds of Christian apostasy. Statistics for the full models of Study 2 are in the Appendix beginning with Table A4.

Mini-replication of analytic apostasy. To quickly gauge the possibility that these analytic apostasy results were endemic to MTurk workers, we pulled in 106 Prolific workers from the United States (after excluding five non-completers) to the abbreviated version

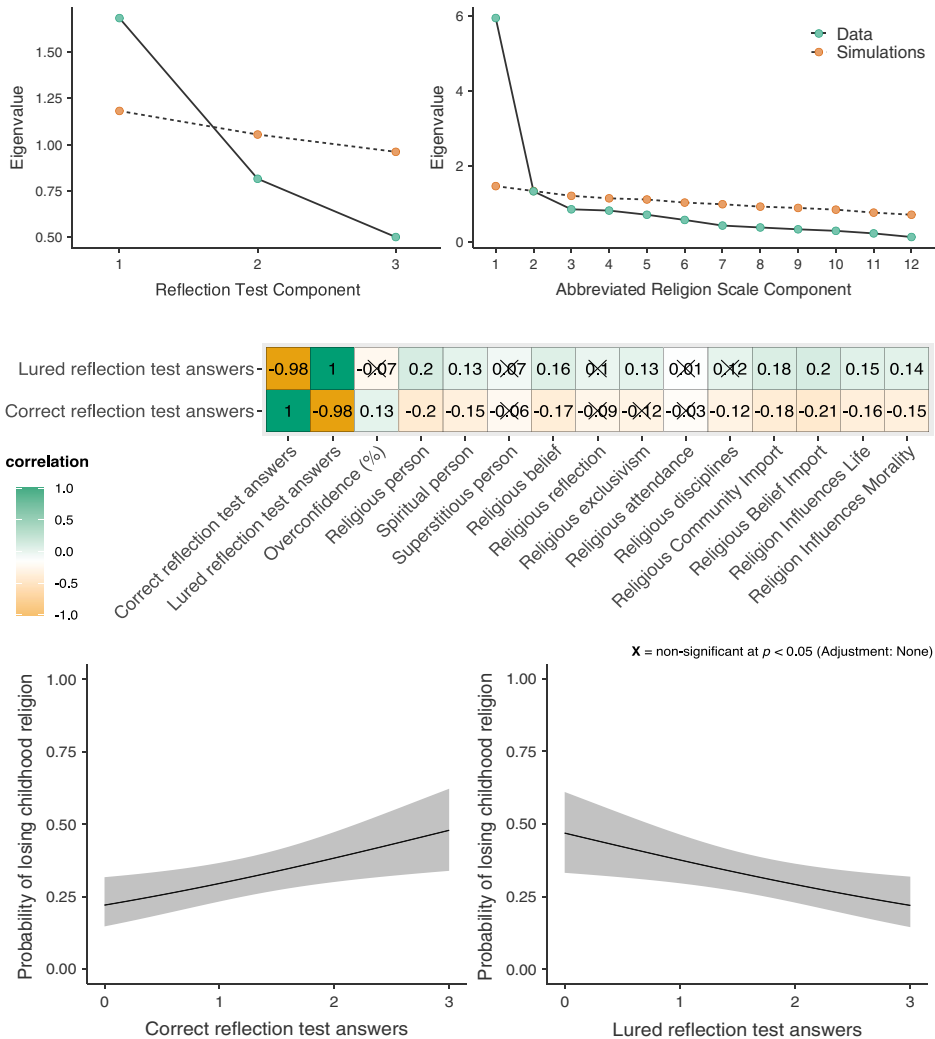


Figure 2. The single principal component for the novel reflection test and abbreviated religion items (top), the bivariate correlations between reflection test answers, religiosity, and spirituality (middle: x means $p \geq 0.05$, holm adjusted) and the multivariate odds of apostasy per correct and lured answer with 95% C.I. (bottom) in study 2.

of this survey. Controlling for the same measures, reflection test performance predicted even more extreme odds of both general apostasy (Correct: OR = 1.95, 95% CI [1.20, 3.16], $p = 0.007$; Lured: OR = 0.52, 95% CI [0.32, 0.86], $p = 0.010$) and Christian apostasy (Correct: OR = 2.23, 95% CI [1.34, 3.72], $p = 0.002$; Lured: OR = 0.44, 95% CI [0.26, 0.75], $p = 0.002$). These relationships remained even when controlling for additional measures in the mini-replication: subjective numeracy, objective numeracy, and attention.

Discussion

Study 2 conceptually replicated and extended analytic atheism. People who had more correct reflection test answers were less religious (on average) than people who had fewer

correct answers. Likewise, the more that people fell for the lure on reflection tests, the more religious they were. Extending those *between* person results are the *within* person *analytic apostasy* results: the probability of a person losing their religion since childhood was predicted by both correct and lured reflection test answers. The better someone performed on the reflection test, the higher their odds of being an apostate and the more that someone fell for the lure on the reflection test question, the lower their odds of being an apostate. Like Study 1, these observational results are considered small, at least by some researchers in Epidemiology (Chen *et al.* 2010). Nonetheless, the successful mini-replication among participants from another source suggests the result is somewhat robust.

Be that as it may, Study 2 had notable limitations. First, all Study 2 participants were from the United States. As such, Study 2 cannot (by itself) support the conclusion that analytic apostasy transcends a particular cultural context. Second, another interesting hypothesis is whether – in addition to religious *deconversion* (apostasy) – reflection predicts religious *conversion* (i.e., becoming religious). That is, does reflection predict changing one's mind about religion in *both* directions? Unfortunately, conversion seems rare: Study 2 had only nine respondents (4 per cent) that indicated religious conversion and the mini-replication had half as many converts. Given this, Study 2 could not detect this hypothetical analytic conversion even if it existed in the target population.

Study 3

Study 3 was a pre-registered replication and extension of the results of Study 2 (<https://osf.io/3dvvgk>). The main goal was to test the replicability of analytic atheism and analytic apostasy in the United States and test how those results depend on Country (United Kingdom versus United States) or Participant source (mTurk versus Prolific) – all while controlling for a few more potentially confounding variables, such as numeracy.

Method

The results of prior studies enabled us to better measure our intended constructs in Study 3, and with fewer yet more cross-cultural measures.

Participants

To maximize the statistical power and efficiency of Study 3, we pulled participants into one survey from the sources of both Study 2 (mTurk) and its mini-replication (Prolific). To allow observational relationships to stabilize (Schönbrodt and Perugini 2013), we aimed to recruit another 250 respondents from each country, per platform. CloudResearch pulled in 265 mTurk workers from the US and Prolific pulled in 528 workers from both the US and the UK. Participants were offered \$1.80 USD. Our analysis excludes only participants who automatically exited the survey after choosing 'I do not consent to participate in this study' on the first page ($n = 34$), did not complete the required portion of the survey ($n = 13$), had a ReCAPTCHA (version 3) score of a likely bot ($n = 3$), or reported a country other than one of the two eligible countries ($n = 2$), leaving a final sample of 741. Descriptive statistics for Study 3 are reported in Table 3.

Materials

Study 3 consolidated the materials of Study 2 to make room for a few more control variables without increasing survey length in a way that could sacrifice data quality (Galesic and Bosnjak 2009).

Table 3. Descriptive statistics for Study 3

	Category	n	Per cent		
Gender	Woman	432	58%		
	Man	294	40%		
	Other	15	2%		
Country	United Kingdom	225	30%		
	United States	516	70%		
	Christian	249	34%		
Religion	Other Religion	80	11%		
	Don't Know	5	1%		
	No Religion	407	54%		
	Mean	Median	SD	Minimum	Maximum
Survey progress	99.97	100	0.31	97	100
Survey duration (seconds)	467.22	398	293.22	135	3031
Survey experience	1.08	1	0.69	−2	2
Attention checks passed	2.99	3	0.07	2	3
ReCAPTCHA (v3) score	0.97	1	0.05	0.7	1
Birthyear	1984.38	1987	12.88	1944	2004
Social conservatism	3.1	3	1.75	1	7
Economic conservatism	3.41	3	1.79	1	7
Childhood unpredictability	6.29	5	3.25	3	15
School years (self)	16.61	17	3.31	0	45
STEM courses	6.3	2	12.77	0	200
Philosophy courses	0.87	0	1.73	0	15
Religiosity (average)	2.62	2.4	1.19	1	5
Spirituality (average)	2.81	2.8	1.16	1	5
Correct test answers	1.83	2	1.12	0	3
Lured test answers	1.11	1	1.09	0	3
Reflectively correct answers	0.48	0	0.73	0	3
Test familiarity	1.57	2	1.22	0	3
Lures considered	1.62	2	1.12	0	3

(Continued)

Table 3. (Continued.)

	Mean	Median	SD	Minimum	Maximum
Objective numeracy	1.58	2	0.64	0	2
Subjective numeracy	6.62	7	2.58	0	10
Math sentiment	0.43	0.5	1.7	−3	3

Study 2 materials. We reused the demographic, education, and reflection test items from Study 2. Other items were removed to streamline the survey. Minor changes or additions to our materials are explained below.

Consolidated religion and spirituality scales. We consolidated our 12-item scale of both religiosity and spirituality scale from Study 2 into two five-item scales: one for religiosity and one for spirituality. The reused items were *person*, *reflection*, *belief import*, *community import*, and one new item was *upbringing* ('I was raised in a [religious/spiritual] household'). These dual five-item scales bought us dissociation between religiosity and spirituality without costing us the data quality of a lengthier survey. Response options ranged from 'Strongly disagree' (1) to 'Strongly agree' (5). All five items from each scale loaded into a single component (factor loadings > 0.6) and reliability was excellent for both scales ($\alpha = 0.86$). So scores on each set of five items were averaged.

Reflection test lure consideration and reflective responding. A substantial minority of people answer reflection tests correctly without actually reflecting (Byrd *et al.* 2023). To reduce such reflection test measurement error, we asked participants that didn't select the lure if they had considered the lure before selecting their answer. Every correct answer that involved consideration of the lure counted as 'reflective'. Reflective answers for the three questions loaded on one component (factor loadings > 0.6, $\alpha = 0.38$) and reflective answers correlated with other reflection and expected political metrics without correlating with gender (Figure A4).

Objective numeracy. Participants completed a die-rolling probability test and a frequency-to-per cent conversion task with the same four-option response format as the reflection test. Correct answers were summed. Scores loaded on the same component (factor loadings > 0.7, $\alpha = 0.4$). We averaged the number of correct answers per participant.

Subjective numeracy. Participants answered both, 'How good are you at figuring out how much something will cost if it is discounted by 1/4?' and, 'How good are you at figuring out how much something will cost if it is discounted by 15%?' on a 6-point scale from 'Not at all good' (0) to 'Extremely good' (5). Scores loaded on the same component (factor loadings > 0.9) with high reliability ($\alpha = 0.82$). Correct answers per respondent were averaged.

Math sentiment. Participants were asked, 'How good are you at doing math?' and, 'How do you feel about math?' on a 7-point scale from 'Terrible' or 'I hate math' (−3) to 'Great' or 'I love math' (3). Scores loaded on the same component (factor loadings > 0.9) with high reliability ($\alpha = 0.87$). This item's score is the average number of correct answers per participant.

Better data quality controls. There is growing evidence that conventional attention checks are either insufficient or counterproductive. For example, the US Centers for Disease Control posted a report that four per cent of survey respondents reported 'drinking or gargling diluted bleach solutions' during the pre-vaccine portion of the COVID - 19 pandemic (Gharpure 2020). CloudResearch replicated this result, but found that '80–90% of reports of household cleanser ingestion' were from respondents who also selected impossible claims such as 'having had a fatal heart attack' or 'eating concrete for its iron content' (Litman *et al.*

2023). CloudResearch has shown that many such low quality responses come from people in developing countries who use virtual private networks and assistance from third parties to feign eligibility for relatively short surveys that pay the equivalent of a day's wages (Moss and Litman 2018). Because we recruited not just from CloudResearch, but also Prolific, we attempted to overcome these data quality issues by adding measures of bot-like behaviour and English proficiency to our instructional attention check from our prior studies (Byrd 2025).

Bot-like behaviour. In addition to the quantitative ReCAPTCHA (v3) metric of bot-like behaviour used in Study 2 (Qualtrics 2022), Study 3 collected qualitative data about respondent effort or comprehension. Participants were shown an image of someone showing a strong negative emotion and given this simple instruction: 'In one complete sentence, explain what may have happened immediately before this photo was taken' (Byrd 2025). Few participants performed poorly on the bot test ($n = 31$). Examples of poor responses contained no English words (e.g., 'Repe'), were not a complete sentence (e.g., 'arrested', 'verbal disagreement'), did not make sense (e.g., 'the girl do fight any one'), or described the photo rather than its antecedent (e.g., 'angry', 'angry woman'). Because even fewer of these respondents remained after the above-mentioned exclusions ($n = 22$), the results do not seem to be impacted by them.

English proficiency. We also asked participants, 'What material is the shirt in the photo above?' Response options included a correct answer 'Cotton' (1) and two words that look similar to someone with poor English proficiency: 'Cobalt' (0) and 'Copper' (0). Correct answers were added to the sum of passed attention checks.

Childhood unpredictability. Between Study 2 and 3, we learned that childhood predictability may be related to both religiosity (Maranges et al. 2021) and reasoning style (Wang et al. 2022). Given our focus on how reasoning style predicts changes in religiosity since childhood, we thought it prudent to measure and control for childhood unpredictability using three items from validated scales. After reading, 'When I was younger than 10,' participants rated their agreement with 'things were often chaotic in my house' (home chaos), 'people often moved in and out of my house on a pretty random basis' (random visitors), and 'I had a hard time knowing what my parent(s) or other people in my house were going to say or do from day-to-day' (unpredictable people) on a 5-point scale from 'Strongly disagree' (1) to 'Strongly agree' (5).

Results

We detected signs of analytic atheism and analytic apostasy using both correct and *reflectively* correct test answers, even when controlling for the other measured variables. Nonetheless, other variables such as participant source and numeracy were also predictive of both reflection test performance and apostasy.

Analytic Atheism. The top of Figure 3 shows how reflection test performance correlated with our religiosity and spirituality items. We detected correlations between half of these religio-spiritual items and lured answers, always positive. We detected similar correlations using the number of lures people considered. Moreover, we found correlations between most of the religio-spiritual items and correct reflection test answers, always negative. Familiarity with the reflection test correlated similarly. We detected religious reflection correlating only with *reflectively* correct answers – and positively: as the number of reflectively correct answers increased, the importance of thinking critically about religion increased. We did not detect correlations between the new religious or spiritual upbringing items and any reflection test metric.

Analytic Apostasy. In Study 3, correct reflection test answers predicted slightly higher odds of general apostasy before controlling for confounds (OR = 1.14, 95% CI [0.99, 1.31],

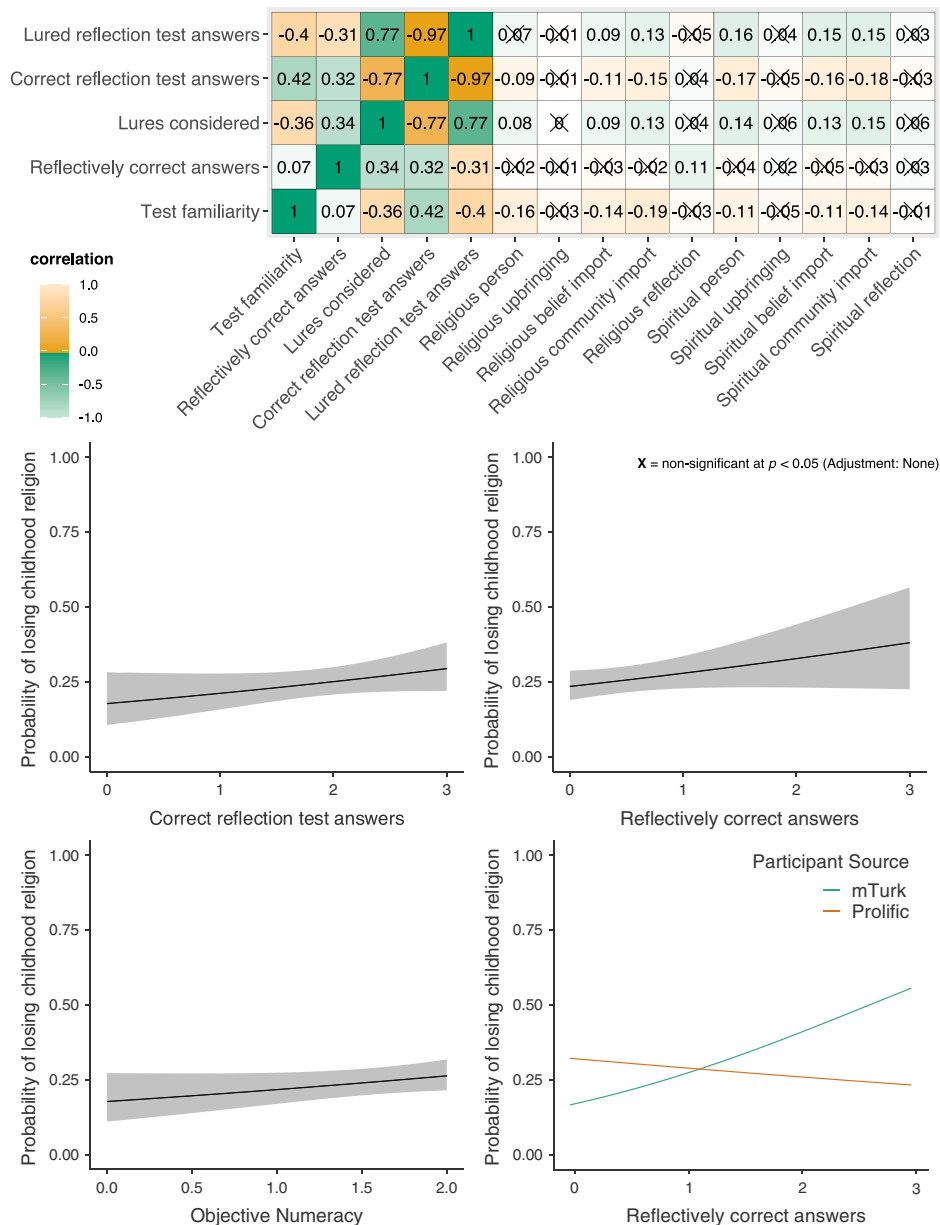


Figure 3. Bivariate correlations between reflection metrics, religion, and spirituality (top; x means $p \geq 0.05$, holm adjusted) and multivariate odds of apostasy by reflection, numeracy, and source with 95% C.I. (bottom) in Study 3.

$p < 0.068$) and – as Figure 3 shows – even higher odds of apostasy when controlling for confounds (OR = 1.55, 95% CI [1.05, 2.27], $p = 0.024$). The higher odds of apostasy attributable to objective numeracy (above and beyond confounds) were more attenuated (OR = 1.29, 95% CI [0.95, 1.75], $p = 0.108$).

Reflectively correct answers also predicted higher odds of general apostasy above and beyond confounds (OR = 1.46, 95% CI [1.01, 2.11], $p = 0.042$). Notably, in this model, the

relationship between reflection and apostasy interacted with participant source: compared to mTurk workers, Prolific workers' reflectively correct answers predicted lower odds of apostasy ($OR = 0.49$, 95% CI $[0.29, 0.83]$, $p = 0.008$). Moreover, in this model Prolific users had more than twice the odds of apostasy compared to the mTurk workers ($OR = 2.29$, 95% CI $[1.39, 3.77]$, $p < 0.001$) even though U.K. participants had nearly half the odds of apostasy as U.S. participants ($OR = 0.57$, 95% CI $[0.34, 0.96]$, $p = 0.035$).

Lured answers did not clearly predict apostasy before controlling for confounds ($OR = 0.89$, 95% CI $[0.77, 1.03]$, $p = 0.113$) or after ($OR = 0.71$, 95% CI $[0.48, 1.05]$, $p = 0.085$).

Contrary to our pre-registered expectations, reflection test performance was not a stronger predictor of Christian apostasy than general apostasy. Odds of Christian apostasy increased with reflectively correct answers, but to a similar degree as general apostasy ($OR = 1.40$), albeit not beyond conventional thresholds of significance (95% CI $[0.96, 2.04]$, $p = 0.095$). Also, neither lured nor merely correct answers were related to Christian apostasy in Study 3 above and beyond the other factors such as country or participant source ($p > 0.177$).

Discussion

Even after improving our measure of reflection, and controlling for more confounds, we replicated signs of analytic atheism and analytic apostasy across countries and participant sources. These replications and extensions strengthen confidence in the results from Study 2.

There are at least three limitations of Study 3 – the first two are the same as the prior study. First, Study 3 sampled from just two Western countries, preventing us from generalizing its results to other countries and cultures. Second, like Study 2, Study 3 was still unable to recruit enough religious converts to test the analytic conversion hypothesis: just 20 people (3 per cent) in Study 3 became religious since childhood. And third, inextricable overlap between country and participant source may make it impossible to disentangle country- or platform-based differences in reflection test familiarity from country- or platform-based differences in analytic atheism or analytic apostasy. The main reason is that all mTurk workers were from the US, while Prolific workers were from either the US or UK. So every country-level analysis is partly a platform-level analysis (and vice versa).

Study 4

To overcome the limitations of Study 3, we needed a larger, more culturally diverse sample. For this we returned to the push method used in our first study.

Method

From February 2023 to March 2024, we pushed an ad to people on English-language Google webpages with English-language browser settings.

Participants

More than 21,000 people entered the survey from the ad and 5,137 completed it. The following analysis excludes only people who reported a country that was not listed in Qualtrics' prepopulated list of 193 countries ($n = 51$).

Materials

People who clicked the ad were directed to a survey that was identical to Study 3, with one difference: Study 4 re-implemented the personality test and score from Study 1 to compensate the ad-recruited participants. Descriptive statistics for Study 4 are reported in [Table 4](#).

Results

In this final push study, we detected signs of analytic atheism and analytic apostasy, but not analytic conversion or analytic spirituality.

Analytic Atheism. [Figure 4](#) (top right) shows how correct reflection test answers predicted lower religiosity ($R = -0.13$, 95% CI $[-0.15, -0.10]$, $p < 0.001$), albeit less so after controlling for confounds ($R = -0.06$, 95% CI $[-0.13, 0.02]$, $p = 0.045$). The full model explained 20 per cent of the variance in religiosity. Replacing correct answers with reflectively correct answers predicted a smaller decrease in religiosity ($R = -0.05$, 95% CI $[-0.08, -0.02]$, $p < 0.001$), but not above and beyond other confounds ($R = -0.06$, $p = 0.225$), such as objective numeracy ($R = -0.05$, 95% CI $[-0.08, -0.03]$, $p < 0.001$). Nonetheless, correct responses predicted apostasy in nearly every United Nations region ([Figure 4](#), bottom right). Only the 55 participants from Oceania bucked the trend.

In the same model, correct answers also predicted lower spirituality ($R = -0.10$, 95% CI $[-0.13, -0.08]$, $p < 0.001$) until controlling for confounds ($R = -0.03$, $p = 0.266$). Relationships between spirituality and *reflectively* correct answers were not detected before or after controlling for confounds ($p > 0.159$).

Analytic Apostasy. Study 4 replicated the prior study's analytic apostasy result: correct reflection test answers predicted greater odds of apostasy (OR = 1.33, 95% CI $[1.23, 1.43]$, $p < 0.001$), even after controlling for confounds (OR = 1.59, 95% CI $[1.03, 2.45]$, $p = 0.035$). Replacing correct answers with *reflectively* correct test answers in this model predicted similarly higher odds of apostasy (OR = 1.31, 95% CI $[1.18, 1.46]$, $p < 0.001$) until controlling for all confounds (OR = 1.42, $p = 0.271$).

The results for Christian apostasy were nearly identical. The odds of *Christian* apostasy were predicted to increase by at least as much for correct reflection answers (OR = 1.36, 95% CI $[1.25, 1.48]$, $p < 0.001$), even after controlling for confounds (OR = 1.71, 95% CI $[1.06, 2.76]$, $p = 0.028$). Replacing correct with *reflectively* correct answers predicted similarly higher odds of Christian apostasy (OR = 1.32, 95% CI $[1.17, 1.49]$, $p < 0.001$) until controlling for confounds (OR = 1.59, $p = 0.190$).

Analytic Conversion. Study 4 recruited enough religious converts to test the analytic conversion hypothesis. However, the odds of converting to religion were not predicted by correct, reflectively correct, or lured answers before or after controlling for confounds ($p > 0.502$). Instead of analytic conversion, we detected that odds of conversion were higher among women (compared to men: OR = 1.59, 95% CI $[1.05, 2.41]$, $p = 0.028$), among people from Europe (compared to the Americas: OR = 2.06, 95% CI $[1.03, 4.10]$, $p = 0.040$), and varied with childhood unpredictability (OR = 1.07, 95% CI $[1.01, 1.13]$, $p = 0.026$), controlling for other factors. Controlling for the same factors, odds of conversion decreased as agreeableness increased (OR = 0.89, 95% CI $[0.84, 0.95]$, $p < 0.001$) and math sentiment increased (OR = 0.81, 95% CI $[0.73, 0.89]$, $p < 0.001$).

Can analytic apostasy explain analytic atheism? The size and measures of Study 4 also allowed us to test whether analytic atheism is explained by analytic apostasy: if removing the apostates from the sample eliminates the analytic atheist correlation, then perhaps only apostates are more reflective and life-long atheists and agnostics are about as reflective as religious believers. Sure enough, when apostates were filtered out of the sample, the

Table 4. Descriptive statistics for Study 4

	Category	n	Per cent		
Gender	Woman	3172	62%		
	Man	1674	33%		
	Other	240	5%		
Region ⁴	Africa	302	6%		
	Americas	1708	34%		
	Asia	1252	25%		
	Europe	1769	35%		
	Oceania	55	1%		
Religion	Christian	1863	37%		
	Other Religion	1390	27%		
	Don't Know	286	6%		
	No Religion	1547	30%		
	Mean	Median	SD	Minimum	Maximum
Survey progress	99.74	100	1.68	89	100
Survey duration (seconds)	849.18	721	532.79	174	10157
Survey experience	0.52	0	0.72	-2	2
Attention checks passed	2.58	3	0.76	0	3
ReCAPTCHA (v3) score	0.94	1	0.08	0.1	1
Birthyear	1998.89	2002	9.73	1900	2005
Social conservatism	3.79	4	1.64	1	7
Economic conservatism	3.92	4	1.5	1	7
Childhood unpredictability	8.25	8	3.06	3	15
School years (self)	14.99	15	4.88	0	99
STEM courses	10.47	2	32.81	0	1056
Philosophy courses	2.3	0	9.45	0	100
Religiosity (average)	15.74	16	5.8	5	25
Spirituality (average)	15.66	16	5.39	5	25
Correct test answers	1.33	1	0.95	0	3
Lured test answers	1.47	2	0.95	0	3
Reflectively correct answers	0.4	0	0.61	0	3

(Continued)

Table 4. (Continued.)

	Mean	Median	SD	Minimum	Maximum
Test familiarity	0.87	1	0.89	0	3
Lures considered	1.97	2	0.92	0	3
Objective Numeracy	1	1	0.74	0	2
Subjective Numeracy	5.82	6	3.04	0	10
Math sentiment	0.63	1	1.8	-3	3

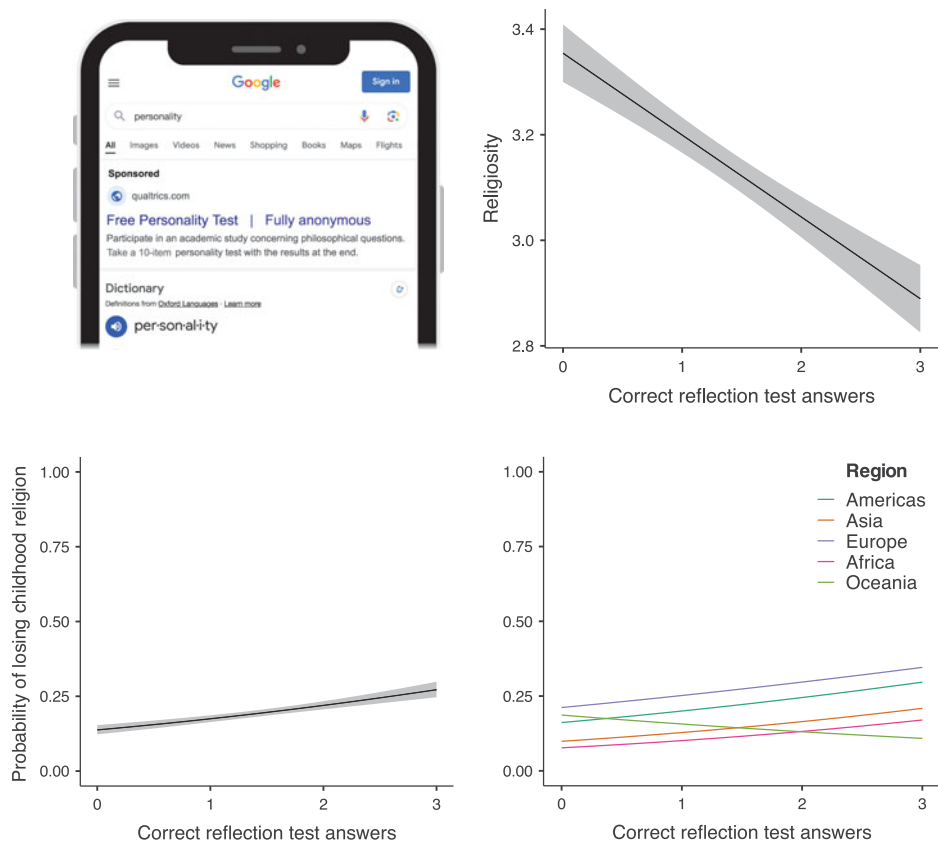


Figure 4. Example ad for recruiting participants (top left), correct reflection test answers predicted lower religiosity (top right) and higher odds of apostasy (bottom left), by region (bottom right).

analytic atheist correlation became non-significant ($n = 4108$, $R = -0.03$, 95% CI $[-0.10, 0.05]$, $p = 0.326$). Moreover, apostates performed better on reflection tests than all others ($r_{correct} = 0.11$, 95% CI $[0.08, 0.13]$, $p < 0.001$; $r_{reflectively\ correct} = 0.07$, 95% CI $[0.04, 0.10]$, $p < 0.001$; $r_{lured} = -0.08$, 95% CI $[-0.06, -0.11]$, $p < 0.001$). Together, these results suggest that analytic atheism may be largely explained by apostasy, not atheism *per se*.

Discussion

Study 4 replicated the analytic atheist, analytic apostasy, and analytic Christian apostasy relationships we pre-registered for Study 3. We also found analytic atheism was largely explained by apostates' outstandingly reflective thinking. However, we did not detect signs of analytic conversion. As such, reflection's relationship to religious changes were not symmetrical: reflection tended to predict deconversion, but not conversion.

General discussion

The latest research continues to find that reflective reasoning predicts disbelief in God (Ghasemi et al. 2024). And even in a meta-analysis in which (atheist) apostates were the most reflective, both converts and apostates who changed most of their beliefs were more reflective than theists who never changed their beliefs (Stagnaro and Pennycook 2025). Our results conceptually replicate these results and extend them with studies that mitigate measurement error, control for more confounds, analyse within-person changes in religiosity, and exploit push recruitment for samples that are larger and more diverse than is typical.

Since cognitive scientists of religion deemed religiosity more intuitive and areligiosity more reflective, there has been disagreement about the normative upshots of analytic atheism. Some have suggested that the intuitiveness of religious belief is a mark against believing, given the limitations of intuition (Guthrie 1993). However, others have argued that the intuitiveness of religious belief is actually a point against disbelief (Barrett and Church 2013). Another view is to resist normative conclusions about the correlates of intuition and reflection, given that demographics are among the correlates, and, therefore, could lead to controversial conclusions (Easton 2018). Of course, some philosophers of religion have argued that religious beliefs are normatively unrelated to intuition or reflection because *religious* beliefs are 'properly basic' and, therefore, do not require reflective thinking (Plantinga 1967; cf. De Cruz 2014). This line of thinking may be popular among ordinary people who often employ more permissive epistemic standards for religion than science (Davoodi and Lombrozo 2022a, 2022b; Liquin et al. 2020; Metz et al. 2023). However, even believing scholars who have seen our results have admitted that because reflection tests predict better judgment in many domains, it is difficult to see how analytic apostasy could be as favourable to religious belief as disbelief.⁵ We look forward to further analysis of analytic atheism and analytic apostasy from our colleagues in religious studies.

Conclusion

Thousands of people from dozens of countries were recruited in multiple ways to complete multiple survey instruments to triangulate on the relationship between religiosity and reflective thinking. They repeatedly exhibited signs of analytic atheism: the more reflectively people reasoned, the less religious they were. Although less reliable, they also exhibited signs of analytic aspirotuality: the more reflectively people reasoned, the less spiritual they were. Finally, they exhibited analytic apostasy: the more reflectively a person reasoned, the higher that person's odds were of losing their religion since childhood. Importantly, our data also suggests that atheists appear more reflective in large part because the apostates in their midst are significantly more reflective than others. The reported relationships were small, which is compatible with cases of exceptionally reflective believers and converts. So analytic atheism, analytic aspirotuality, and analytic apostasy may describe phenomena that occur at the margins (in the economist's sense): reflective thinking does not guarantee or even characterize areligiosity or apostasy, but (on average)

reflective thinking does seem to predict more decreases in religiosity than increases – not just between individuals, but within individuals.

Acknowledgements. Helpful suggestions and ideas were provided by Justin Barrett, Gina Bolton, Ian Church, Helen De Cruz, Johan De Smedt, John Horgan, Joshua Knobe, Tamar Kushnir, Michael Prinzing, Blake McAllister, Edouard Machery, Jennifer McBryan, Ameni Mehrez, Ryan Nichols, Shaun Nichols, Paul Rezkalla, and Jim Spiegel.

Financial support. This research was supported by the John Templeton Foundation (61886).

Author contributions. CRediT taxonomy (<http://credit.niso.org>). Conceptualization: NB, SS, JS; Data curation: NB, JS; Formal analysis: NB, JS; Funding acquisition: NB, SS, JS; Investigation: NB, SS, JS; Methodology: NB, SS, JS; Project Administration: NB, SS, JS; Visualization: NB; Writing – original draft: NB; Writing – review & editing: NB, SS, JS.

Notes

1. Because women have reported different levels of religiosity and performed differently on reflection tests in prior research, we have included gender in our analyses. When fewer than 50 respondents reported a gender besides ‘Man’ or ‘Woman’, the remaining gender category has insufficient statistical power to predict variance in any other variable. So for our studies that collected so few non-binary responses, our multivariate inferential statistical models collapsed ‘Man’ and ‘Other’ into a single category. This optimized statistical power, included otherwise marginalized people, and controlled for the potentially confounding factor of reporting a gender of ‘Woman’. For transparency, our Jamovi analyses for these studies also include the version of the models with all three gender categories. The overall pattern of results is the same in those analyses, albeit with less statistical power due to less inclusion.
2. We also piloted a diverse set of five new reflection test items on about 50 participants, but the base rate neglect, intuitive physics, and belief bias adaptations did not load onto the same component as the three that all Study 2 participants completed. Also, the remaining Wason selection and conjunction fallacy items did not load as strongly on the same component as the three that all Study 2 participants completed. Finally, the internal reliability of the additional five items was not better than the prior three. So subsequent analyses are based on the three new reflection test items that all Study 2 participants completed.
3. We did not include both correct and lured answers or any interactions in the multivariate analysis for Study 2 because this resulted in variance inflation factors (VIF) above five, introducing a potential multicollinearity issue in such a small sample.
4. Most countries in the dataset had too few respondents, preventing their country from predicting variance in other observed variables. To include these respondents data without sacrificing statistical power, Qualtrics’s country codes were collapsed into one of six continental or ‘geographic regions’ according to the United Nations geoscheme (Statistics Division 2021).
5. In addition to normative implications about religious *belief*, some believers have pointed out that results like ours may generate interesting hypotheses about religious *doctrine* and *practice*. For example, some have argued that analytic atheism is compatible with some religious groups being more inclusive regarding cognitive ability than some non-religious groups (Spiegel 2015). The idea seems to be that if certain religious groups are more accepting of people with lower cognitive ability than non-religious groups, then you would expect to find that aggregate cognitive test performance is lower among religious groups than non-religious groups. Although our data cannot logically support the hypothesis by affirming its consequent, we invite research that could validly find support for the hypothesis. Further, we hope research can assess how well this hypothesis fits our detection of analytic *apostasy* or our failure to detect analytic *conversion*.

References

- Altemeyer B and Hunsberger B (2004) A Revised Religious Fundamentalism Scale: The Short and Sweet of It. *The International Journal for the Psychology of Religion* **14**(1), 47–54. https://doi.org/10.1207/s15327582ijpr1401_4
- Antoun C, Zhang C, Conrad FG and Schober MF (2016) Comparisons of Online Recruitment Strategies for Convenience Samples: Craigslist, Google AdWords, Facebook, and Amazon Mechanical Turk. *Field Methods* **28**(3), 231–246. <https://doi.org/10.1177/1525822X15603149>
- Attali Y and Bar-Hillel M (2020) The False Allure of Fast Lures. *Judgment and Decision Making* **15**(1), 93–111. <https://doi.org/10.1017/S1930297500006938>

- Barrett JL and Church IM (2013) Should CSR Give Atheists Epistemic Assurance? On Beer-Goggles, BFFs, and Skepticism Regarding Religious Beliefs. *The Monist* 96(3), 311–324. <https://doi.org/10.5840/monist201396314>
- Batson CD and Schoenrade PA (1991) Measuring Religion as Quest: 2) Reliability Concerns. *Journal for the Scientific Study of Religion* 30(4), 430–447. <https://doi.org/10.2307/1387278>
- Byrd N (2022) Bounded Reflectivism & Epistemic Identity. *Metaphilosophy* 53(1), 53–69. <https://doi.org/10.1111/meta.12534>
- Byrd N (2023) Great Minds do not Think Alike: Philosophers' Views Predicted by Reflection, Education, Personality, and Other Demographic Differences. *Review of Philosophy and Psychology* 14(2), 647–684. <https://doi.org/10.1007/s13164-022-00628-y>
- Byrd N (2019) All Measures Are Not Created Equal: Reflection Test, Think Aloud, and Process Dissociation Protocols. <https://researchgate.net/publication/344207716> (accessed 15 March 2025)
- Byrd N (2025) Reflection-Philosophy Order Effects and Correlations Across Samples. *Analysis*. <https://doi.org/10.1093/analys/anaf015>
- Byrd N and Conway P (2019) Not All Who Ponder Count Costs: Arithmetic Reflection Predicts Utilitarian Tendencies, But Logical Reflection Predicts Both Deontological and Utilitarian Tendencies. *Cognition*, 192. <https://doi.org/10.1016/j.cognition.2019.06.007>
- Byrd N, Joseph B, Gongora G and Sirota M (2023) Tell Us What You Really Think: A Think Aloud Protocol Analysis of the Verbal Cognitive Reflection Test. *Journal of Intelligence* 11(4). <https://doi.org/10.3390/jintelligence11040076>
- Calvillo DP, Bratton J, Velazquez V, Smelter TJ and Crum D (2023) Elaborative Feedback and Instruction Improve Cognitive Reflection But Do Not Transfer to Related Tasks. *Thinking & Reasoning* 29(2), 276–304. <https://doi.org/10.1080/13546783.2022.2075035>
- Chen H, Cohen P and Chen S (2010) How Big is a Big Odds Ratio? Interpreting the Magnitudes of Odds Ratios in Epidemiological Studies. *Communications in Statistics - Simulation and Computation* 39(4), 860–864. <https://doi.org/10.1080/03610911003650383>
- Davoodi T and Lombrozo T (2022a) Explaining the Existential: Scientific and Religious Explanations Play Different Functional Roles. *Journal of Experimental Psychology: General* 151(5), 1199–1218. <https://doi.org/10.1037/xge0001129>
- Davoodi T and Lombrozo T (2022b) Varieties of Ignorance: Mystery and the Unknown in Science and Religion. *Cognitive Science* 46(4), e13129. <https://doi.org/10.1111/cogs.13129>
- De Cruz H (2014) The Enduring Appeal of Natural Theological Arguments. *Philosophy Compass* 9(2), 145–153. <https://doi.org/10.1111/phc3.12105>
- Easton C (2018). Women and 'the Philosophical Personality': Evaluating Whether Gender Differences in the Cognitive Reflection Test Have Significance for Explaining the Gender Gap in Philosophy. *Synthese* (October, 31). <https://doi.org/10.1007/s11229-018-01986-w>
- Erceg N, Galic Z, and Ružojčić M (2020) A Reflection on Cognitive Reflection – Testing Convergent Validity of Two Versions of the Cognitive Reflection Test. *Judgment & Decision Making* 15(5), 741–755. <https://doi.org/10.1017/S1930297500007907>
- Feltz A and Cokely ET (2011) Individual Differences in Theory-of-Mind Judgments: Order Effects and Side Effects. *Philosophical Psychology* 24(3), 343–355. <https://doi.org/10.1080/09515089.2011.556611>
- Fisher AJ, Medaglia JD and Jeronimus BF (2018) Lack of Group-to-Individual Generalizability is a Threat to Human Subjects Research. *Proceedings of the National Academy of Sciences* 115(27), E6106–E6115. <https://doi.org/10.1073/pnas.1711978115>
- Frederick S (2005) Cognitive Reflection and Decision Making. *Journal of Economic Perspectives* 19(4), 25–42. <https://doi.org/10.1257/089533005775196732>
- Galesic M and Bosnjak M (2009) Effects of Questionnaire Length on Participation and Indicators of Response Quality in a Web Survey. *Public Opinion Quarterly* 73(2), 349–360. <https://doi.org/10.1093/poq/nfp031>
- Gervais WM, van Elk M, Xygalatas D, McKay RT, Aveyard M, Buchtel EE, Dar-Nimrod I, Klocová EK, et al. (2018) Analytic Atheism: A Cross-culturally Weak and Fickle Phenomenon? *Judgment and Decision Making* 13(3), 268–274. <https://doi.org/10.1017/S1930297500007701>
- Gette CR, & Kryjevskaja M (2019) Establishing a relationship between student cognitive reflection skills and performance on physics questions that elicit strong intuitive responses. *Physical Review Physics Education Research*, 15. <https://doi.org/10.1103/PhysRevPhysEducRes.15.010118>
- Gharpure R (2020) Knowledge and Practices Regarding Safe Household Cleaning and Disinfection for COVID-19 Prevention — United States, May 2020. *MMWR Morbidity and Mortality Weekly Report*, 69. <https://doi.org/10.15585/mmwr.mm6923e2>
- Ghasemi O, Yilmaz O, Isler O, Terry J and Ross RM (2024) Reflective Thinking Predicts Disbelief in God across 19 Countries. Open Science Framework. <https://doi.org/10.31234/osf.io/wny9p>

- Goplen J and Plant EA (2015) A Religious Worldview Protecting One's Meaning System Through Religious Prejudice. *Personality and Social Psychology Bulletin* **41**(11), 1474–1487. <https://doi.org/10.1177/0146167215599761>
- Guthrie SE (1993) *Faces in the Clouds: A New Theory of Religion*. New York: Oxford University Press.
- Hauser DJ, Moss AJ, Rosenzweig C, Jaffe SN, Robinson J and Litman L (2023) Evaluating CloudResearch's Approved Group as a Solution for Problematic Data Quality on MTurk. *Behavior Research Methods* **55**, 3953–3964. <https://doi.org/10.3758/s13428-022-01999-x>
- Hodge DR (2003) The Intrinsic Spirituality Scale. *Journal of Social Service Research* **30**(1), 41–61. https://doi.org/10.1300/J079v30n01_03
- Korsgaard CM (1996) *The Sources of Normativity*. Cambridge: Cambridge University Press.
- Kung FYH, Kwok N and Brown DJ (2018) Are Attention Check Questions a Threat to Scale Validity? *Applied Psychology* **67**(2), 264–283. <https://doi.org/10.1111/apps.12108>
- Levine S, Rottman J, Davis T, O'Neill E, Stich S and Machery E (2021) Religious Affiliation and Conceptions of the Moral Domain. *Social Cognition* **39**(1), 139–165. <https://doi.org/10.1521/soco.2021.39.1.139>
- Liberali JM, Reyna VF, Furlan S, Stein LM and Pardo ST (2012) Individual Differences in Numeracy and Cognitive Reflection, with Implications for Biases and Fallacies in Probability Judgment. *Journal of Behavioral Decision Making* **25**(4), 361–381. <https://doi.org/10.1002/bdm.752>
- Liquin EG, Metz SE and Lombrozo T (2020) Science Demands Explanation, Religion Tolerates Mystery. *Cognition*, 204. <https://doi.org/10.1016/j.cognition.2020.104398>
- Litman L, Rosen Z, Hartman R, Rosenzweig C, Weinberger-Litman SL, Moss AJ and Robinson J (2023) Did People Really Drink Bleach to Prevent COVID-19? A Guide for Protecting Survey Data Against Problematic Respondents. *PLOS ONE* **18**(7). <https://doi.org/10.1371/journal.pone.0287837>
- Livengood J, Sytsma J, Feltz A, Scheines R and Machery E (2010) Philosophical Temperament. *Philosophical Psychology* **23**(3), 313–330. <https://doi.org/10.1080/09515089.2010.490941>
- Machery E, Stich S, Rose D, Chatterjee A, Karasawa K, Struchiner N, Sirker S, Usui N, et al. (2017) Gettier Across Cultures. *Noûs* **51**(3), 645–664. <https://doi.org/10.1111/nous.12110>
- Maranges HM, Hasty CR, Maner JK and Conway P (2021) The Behavioral Ecology of Moral Dilemmas: Childhood Unpredictability, But Not Harshness, Predicts Less Deontological and Utilitarian Responding. *Journal of Personality and Social Psychology* **120**(6), 1696–1719. <https://doi.org/10.1037/pspp0000368>
- Metz SE, Liquin EG and Lombrozo T (2023) Distinct Profiles for Beliefs About Religion Versus Science. *Cognitive Science* **47**(11). <https://doi.org/10.1111/cogs.13370>
- Moss A and Litman L (2018) After the Bot Scare: Understanding What's Been Happening with Data Collection on MTurk and How to Stop It. <https://cloudresearch.com/resources/blog/after-the-bot-scare-understanding-whats-been-happening-with-data-collection-on-mturk-and-how-to-stop-it/>
- Murray D, Sytsma J and Livengood J (2013) God Knows (But Does God Believe?). *Philosophical Studies* **166**(1), 83–107. <https://doi.org/10.1007/s11098-012-0022-5>
- Norenzayan A and Gervais WM (2012) The Origins of Religious Disbelief. *Trends in Cognitive Sciences* **17**(1), 20–25. <https://doi.org/10.1016/j.tics.2012.11.006>
- Pennycook G, Ross RM, Koehler DJ and Fugelsang JA (2016) Atheists and Agnostics Are More Reflective than Religious Believers: Four Empirical Studies and a Meta-Analysis. *PLOS ONE* **11**(4). <https://doi.org/10.1371/journal.pone.0153039>
- Pigden C 2013. Analytic Philosophy. In Sullivant S and Ruse M (eds.). *The Oxford Handbook of Atheism*. Oxford: Oxford University Press: 307–319.
- Plantinga A (1967) *God and Other Minds: A Study of the Rational Justification of Belief in God*. Ithaca: Cornell University Press.
- Qualtrics. 2022. *Fraud Detection*. <https://www.qualtrics.com/support/survey-platform/survey-module/survey-checker/fraud-detection/> (accessed 15 June 2022).
- Saribay SA and Yilmaz O (2018) Relationships between Core Ideological Motives, Social and Economic Conservatism, and Religiosity: Evidence from a Turkish Sample. *Asian Journal of Social Psychology* **21**(3), 205–211. <https://doi.org/10.1111/ajsp.12213>
- Saribay SA, Yilmaz O and Körpe GG (2020) Does Intuitive Mindset Influence Belief in God? A Registered Replication of Shenhav, Rand and Greene. *Judgment and Decision Making* **15**(2), 193–202. <https://doi.org/10.1017/S1930297500007348>
- Schönbrodt FD and Perugini M (2013) At What Sample Size Do Correlations Stabilize? *Journal of Research in Personality* **47**(5), 609–612. <https://doi.org/10.1016/j.jrp.2013.05.009>
- Shenhav A, Rand DG and Greene JD (2012) Divine Intuition: Cognitive Style Influences Belief in God. *Journal of Experimental Psychology: General* **141**(3), 423–428. <https://doi.org/10.1037/a0025391>

- Sirota M and Juanchich M (2018) Effect of Response Format on Cognitive Reflection: Validating a Two- and Four-Option Multiple Choice Question Version of the Cognitive Reflection Test. *Behavior Research Methods* **50**(6), 2511–2522. <https://doi.org/10.3758/s13428-018-1029-4>
- Sosa E (2009) *Reflective Knowledge: Apt Belief and Reflective Knowledge*. New York: Oxford University Press.
- Spiegel JS (2015) Dumb Sheep by James S. Spiegel. *Touchstone: A Journal of Mere Christianity* June. <https://www.touchstonemag.com/archives/article.php?id=28-03-020-v> (accessed 13 May 2024).
- Stagnaro M, and Pennycook G (2025) On the Role of Analytic Thinking in Religious Belief Change: Evidence from over 50,000 Participants in 16 Countries *Cognition* 254. <https://doi.org/10.1016/j.cognition.2024.105989>
- Statistics Division (2021) *Standard Country or Area Codes for Statistical Use*. United Nations. <https://unstats.un.org/unsd/methodology/m49/> (accessed 01 March 2024).
- Tiliopoulos N, Bikker AP, Coxon APM and Hawkin PK (2006) The Means and Ends of Religiosity: A Fresh Look at Gordon Allport's Religious Orientation Dimensions. *Personality and Individual Differences* **42**(8), 1609–1620. <https://doi.org/10.1016/j.paid.2006.10.034>
- Wang X, Zhu N and Chang L (2022) Childhood Unpredictability, Life History, and Intuitive versus Deliberate Cognitive Styles. *Personality and Individual Differences* **184**. <https://doi.org/10.1016/j.paid.2021.111225>
- Yilmaz O, Adil Saribay S and Iyer R (2020) Are Neo-liberals More Intuitive? Undetected Libertarians Confound the Relation between Analytic Cognitive Style and Economic Conservatism. *Current Psychology* **39**(1), 25–32. <https://doi.org/10.1007/s12144-019-0130-x>
- Yilmaz O and Isler O (2019) Reflection Increases Belief in God through Self-Questioning among Non-Believers. *Judgment and Decision Making* **14**(6), 649–657. <https://doi.org/10.1017/S1930297500005374>

