



This is a repository copy of *Measuring religious evil: examining the use of single-item measures in religious research*.

White Rose Research Online URL for this paper:  
<https://eprints.whiterose.ac.uk/219303/>

Version: Published Version

---

**Article:**

Clark, T. [orcid.org/0000-0001-6871-629X](https://orcid.org/0000-0001-6871-629X), Desmond, S.A. and Bader, C.D. (2024)  
Measuring religious evil: examining the use of single-item measures in religious research.  
Review of Religious Research. ISSN 0034-673X

<https://doi.org/10.1177/0034673x241279424>

---

**Reuse**

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:  
<https://creativecommons.org/licenses/>

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.



[eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk)  
<https://eprints.whiterose.ac.uk/>

# Measuring Religious Evil: Examining the Use of Single-Item Measures in Religious Research

*Review of Religious Research*

1–30

© Religious Research Association 2023



Article reuse guidelines:

[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)

DOI: 10.1177/0034673X241279424

[journals.sagepub.com/home/rrr](https://journals.sagepub.com/home/rrr)

Tom Clark<sup>1</sup> , Scott A. Desmond<sup>2</sup> and  
Christopher D. Bader<sup>3</sup>

## Abstract

Multiple-item indexes are ubiquitous within the sociology of religion. However, there are a growing number of articles in other disciplines that have advocated the use of single-item measures in specific circumstances. Using quantitative survey data taken from the United Kingdom, this paper contributes to this literature by exploring the impact of single and multiple item measures of religious evil on a series of social and political attitudes. The findings suggest that belief in the devil is the most consistent predictor within a multiple-item measure of religious evil and the multiple-item measure does not significantly outperform single-item measures. Indeed, the item “most evil in the world is caused by the devil” could be a more efficient measure of religious evil, particularly where it is combined with religious attendance. While further cross-cultural research on the impact of belief in religious evil remains necessary, the paper also finds some evidence to suggest that exploration of more secular beliefs in evil might be advantageous.

## Keywords

religion, evil, Satan, demons, hell, single-item measures, multiple-item measures

---

<sup>1</sup>The University of Sheffield, UK

<sup>2</sup>Indiana University—Columbus, USA

<sup>3</sup>Chapman University, Orange County, CA, USA

## Corresponding author:

Tom Clark, Department of Sociological Studies, The University of Sheffield, The Wave, Whitman Road, Sheffield S10 2TU, UK.

Email: [t.clark@sheffield.ac.uk](mailto:t.clark@sheffield.ac.uk)

## **Introduction**

It is somewhat of a truism within survey research to suggest that multiple-item measures (MIMs) are more reliable and more valid than single-item measures (SIMs). There are at least three arguments typically made in favor of MIMs. The first is that an assessment of internal reliability cannot be made with a single item. Second, multiple items are less vulnerable to random measurement error, with single items often viewed as being susceptible to biases in meaning and interpretation. Thirdly, multiple items can capture a greater range of dimensions associated with a construct than a single measure, which are presumed to have lower content validity.

However, there are several good reasons to use single items in specific circumstances. From a theoretical point of view, some measures can be considered “doubly concrete” in that they are unambiguous in meaning and attribution (Bergkvist and Rossiter 2007). That is to say that some concepts tend toward binary in terms of interpretation and feeling. Not everything is a multifaceted concept and, in some instances, increasing measurement may contribute to additional response error. Similarly, there are practical advantages to single-item measures, particularly in terms of cost and attrition. This is especially the case for survey research where the accuracy and efficiency of survey items can be at a premium. Not only are single-item measures less monotonous and less time-consuming for respondents, they are also more cost-effective for researchers and funders. Single-item measures may also reduce “method-effects” by reducing the chance of common method bias (Gardner et al. 1998). This typically occurs when relationships between observations occur due to the format of a survey battery rather than the content of items.

These advantages have been noted in many disciplines. Within the field of organizational sciences, for example, Matthews, Pineault, and Hong (2022) recently reported a series of studies that collectively demonstrated that 75 out of the 91 single-item constructs they examined had very good or extensive validity in comparison to their multiple-item counterparts. They concluded that SIMs could: evidence moderate to high content validity; have no usability concerns; demonstrate moderate to high test-retest reliability; and, have extensive criterion validity. Elsewhere, Bergkvist (2016) has made very strong and repeated arguments for the use of single-item measures in marketing research (Bergkvist 2015; Bergkvist and Rossiter 2007, 2009), and there are also an increasing number of papers in the psychological sciences that have demonstrated the utility of SIMs (Allen, Iliescu, and Greiff 2022).

Given the existing research in other disciplines, in this paper we highlight the differences between single-item measures and multiple-item measures and implement methods that can be used to compare their effectiveness. Although many concepts in the sociology of religion are measured by combining multiple items, for this paper we chose to focus on the different approaches that people have taken to measure belief in religious evil. There is a wealth of evidence to demonstrate the association between belief in religious evil and a variety of political opinions and moral attitudes. In most of these studies, belief in religious evil is assessed with a MIM combining three items: belief in the devil, belief in hell, and belief in demons.

However, more recent evidence provided by Desmond, Clark, and Bader (2023) suggests that a single measure of belief in the Devil/Satan can be just as useful in predicting attitudes related to abortion, family matters, sexuality, and substance use—particularly when it is assessed in relation to church attendance. As noted by Baker, Molle, and Bader (2020), images of Satan are relatively uniform—Satan is a powerful, supernatural force of evil that is in opposition to what is considered good (i.e., God). Therefore, belief in the devil can be considered a “doubly concrete” measure in that there is relatively little ambiguity in terms of what that belief connotes.

More generally, the interaction between belief in Satan and church attendance also resonates with the literature on what might be termed “the sociology of evil,” which repeatedly argues for the absolute interdependence of good and evil (Alexander 2001; Douglas 2017; Lemert 1997; Wolff 1969). Against Platonic and Augustinian thinking that would imagine evil as absence, for example, Alexander (2001) argues that what is seen to be evil is always in opposition to the pursuit of what is perceived to be good—or how people think the world ought to be. The investigation of one necessarily entails the investigation of the other. In this respect, Desmond et al. (2023) also demonstrate a significant interaction effect between belief in the Devil/Satan (evil) and church attendance (good) for 10 out of 12 moral beliefs. That is to say that religious service attendance has little or no effect on moral beliefs, unless it is accompanied by a belief in the Devil/Satan.

This paper further explores the utility of a single-item measure of religious evil—belief in the Devil/Satan. More specifically, if a multiple-item measure of religious evil can discriminate between the relative strength of belief because it has “more information,” it is possible to make three predictions. Firstly, the multiple-item measure should be able to account for a greater proportion of the variance in a criterion measure than each of its individual component parts. Second, each component should have the capacity to make a significant contribution to the variance explained. Thirdly, the MIM should also demonstrate the capacity to explain more variance in a criterion measure than other single-item measures of a similar nature. If the multiple-item measure fails to consistently outperform the single-item measure in these respects, then it may be concluded that the single-item measure is just as effective as its multiple-item counterpart.

Using novel survey data collected in the UK, we provide evidence to suggest: a) belief in the devil appears to be the most consistent predictor within the MIM of religious evil; b) the belief in hell and belief in demons components are limited in terms of their additional utility and may even provide a source of measurement variance; c) when combined with a measure of religious attendance, the MIM does not significantly outperform a SIM of belief in the devil. This paper presents evidence to suggest that the item “most evil in the world is caused by the devil” could be a more efficient measure of religious evil, particularly where it is combined with religious attendance. Finally, the paper offers some evidence to suggest that belief in evil may not necessarily be located entirely within supernatural belief systems.

## Single-Item Versus Multiple-Item Measures

There are three main arguments for the use of multiple item measures (MIMs). The first relates to “internal consistency” and suggests that without different components of measurement, it is not possible to estimate reliability. As highlighted by Cho and Kim (2015), Cronbach’s alpha is by far the most used reliability coefficient—with alpha  $\geq .7$  generally considered to be acceptable. When used in this way, alpha is often tacitly taken to be equivalent to a measure of homogeneity, interrelatedness, or general factor saturation—none of which can be assessed by a single item.

Secondly, given that there is no way to assess the reliability of a SIM, there is also no way of telling whether the measurement is subject to more or less random error. It is assumed that increasing the number of items will decrease the likelihood of measurement error. This assumption is commonly based on what is termed “the Spearman-Brown prophecy”—which predicts that measurement errors cancel themselves out as the number of items increase. Both classical test theory and item response theory imply that measurement error decreases as the number of items increases, although the relationship is not necessarily monotonic. While the hard version of such an argument is that the error of a single item is necessarily higher, and the softer version that the reliability of SIMs remains unknown, the essence of the argument is the same: because single items cannot be compared to any corresponding items, SIMs are more problematic than MIMs.

The third argument concerns the multidimensional nature of concepts and their measurement. This is sometimes referred to as “content validity,” and it has two components: concept and measurement. In terms of concept, SIMs cannot capture the complexity of social constructs in the same way that MIMs are able to because they are necessarily limited by number. In terms of measurement, given SIMs typically capture “less information”—often a three, five, or seven, point scale—they are also unable to make sensitive distinctions between respondents. There is some verisimilitude in respect to these issues. In respect to concept, there is little doubt that sophisticated social concepts are unlikely to be suitable for single item measurement—and overly complex questions are usually considered bad practice in survey research. With respect to measurement, SIMs are likely to have fewer response categories and there is little evidence that simply increasing the size of the scale improves measurement by itself (Dawes 2008). Therefore, MIMs appear to be a useful way of managing these issues.

While there were a few authors who questioned the orthodoxy of MIMs (Gorsuch and McFarland 1972; Gorsuch and McPherson 1989; Ray 1974; Wanous and Reichers 1996), it was not until the early 2000s that a growing number of articles began to advocate for the use of single item measures more strongly (Fuchs and Diamantopoulos 2009; Loo 2002; Rossiter 2002). There are two practical, and two technical arguments for the use of SIMs made within this literature. With regards to practicality, SIMs are administratively expedient in terms of time and effort. This is particularly important in survey research where the financial cost of the survey is proportional to the number of items employed—*increase the number*

of items and the cost of the survey will also go up. But in addition to keeping costs down, the extra space that is generated can also be used to include a larger number of constructs. That is to say that SIMs offer better “value for money” than their multiple-item counterparts.

As highlighted by Allen et al. (2022), there are similar benefits from the perspective of the respondent too. Not only do SIMs avoid repetitive content and thereby encourage participation, they might also be preferential in circumstances where cognitive or emotional considerations need to be taken into account, or where time is likely to be a factor in response. It is also true that there are ethical obligations that would dictate that researchers should not needlessly waste the time of respondents with lengthy and repetitious questions.

The technical arguments for the use of SIMs are, perhaps, more convincing from an empirical perspective. The first argument for the use of SIMs in specific cases concerns what Bergkvist and Rossiter (2007, 2009) term “doubly concrete” constructs. These cases occur where the construct of interest is unambiguous, narrow in scope, and where increasing the number of items increases the risk of construct contamination.

Contamination typically has two sources. The first type of contamination occurs where the dimensionality of a construct is extended beyond its focus to include items that are indirectly associated with the target construct, rather than a direct property of it. The second risk of contamination results from what is often referred to as common method bias. This occurs when variance assumed to result from the construct is attributable to the method of measurement. This risk is thought to be particularly prescient when collecting large amounts of data in a single wave. The longer the survey, the more chance that respondents begin to respond to the demands of structure of the battery or instrument rather than the specific requirement of the item.

“Doubly concrete” constructs are, on the other hand, relatively unambiguous in meaning and attribution. This means that adding items for measurement will not provide further dimensionality to the construct because there is no further dimensionality to add. Therefore, any further specification is likely to increase the likelihood of contamination rather than reduce measurement error.

Similarly, the issue of reliability and/or internal consistency is not a concern in such cases. This is because there are no facets of a measure to assess beyond the initial object. The need for a reliability coefficient is negligible. Instead, Bergkvist and Rossiter (2007) suggest using the predictive validity of a criterion to evaluate the performance of SIMs and MIMs. They propose three reasons for this. Firstly, a measure cannot be predictively valid without being reliable (Gorsuch and McFarland 1972); secondly, predictive validity is a more robust measure than content validity; and thirdly, predictive validity is more important than internal consistency (Cho and Kim 2015; Cronbach 1951). If a precise measure is at least comparable in performance with a multiple counterpart, then there is a strong case for the single item approach to be retained.

None of this is to say that SIMs are automatically better than MIMs, but instead the discussion offered here highlights that there are situations where single items

are likely to increase precision rather than necessarily decrease it. One such situation may occur in the measurement of religious evil.

### *The Measurement of Religious Evil*

There is now an emerging body of empirical literature in the USA that has sought to examine the extent of belief in religious evil, and its potential impact on several moral and social issues. Both Grasmick and McGill (1994) and Leiber and Woodrick (1997) used belief in the devil as part of a MIM designed to assess biblical literalism and its various impacts on punitive ideology. Wilcox, Linzey, and Jelen (1991) also developed a four item MIM to examine belief in the threat of Satan to the US ( $\alpha = .69$ ) and used it to explore the political activism of premillennialists—with belief in an active Devil increasing positive attitudes toward activism but decreasing the perceived efficacy of that participation.

Using an eight item “Belief in an Active Satan Scale” ( $\alpha = .81$ ), Wilson and Huff (2001) first demonstrated that belief in religious evil was correlated with negative attitudes about ethnic and sexual minorities. They would later revise the scale to include 10 items ( $\alpha = .91$ ) and demonstrate that such intolerance extended to communists and women, although the relationship was moderated by religious fundamentalism and right-wing authoritarianism (Wilson, Accord, and Bernas 2006).

As highlighted by Baker (2008), while these early studies used aspects of religious evil to categorize particular types of (Christian) religious belief, the work of Wilson and Huff (2001) established that it was possible to measure belief in religious evil, and that such beliefs could influence social and political attitudes. However, it was the first wave of the Baylor Religion Survey (BRS) in 2005 that provided the platform for a more purposeful examination of religious evil (Bader, Mencken, and Froese 2007; Bader et al. 2023). The survey was a nationally representative examination of religious values, practices, and behaviors present in the US population. It contained nearly 400 items devoted to the sociology of religion (Bader et al. 2007). While the early studies of religious evil were relatively idiosyncratic in terms of item construction, the BRS included a standard battery addressing general religious beliefs: “In your opinion does each of the following exist”—“The Devil/Satan,” “Heaven,” “Hell,” “Purgatory,” “Armageddon,” “Angels,” “Demons,” and “The Rapture” (“Absolutely,” “Probably,” “Probably not,” “Absolutely not”). This question set was repeated in subsequent editions of the survey taken in 2007, 2010, and 2014.

Taking three of these items—the Devil/Satan, hell, and demons, Baker (2008) demonstrated the association between each individual item and a range of demographic variables. As income and education levels increase, belief in each item decreases, with women and African Americans tending to hold firmer belief in the Devil, hell, and demons. However, using an additive index of the three items, Baker’s (2008) paper also demonstrated that for women belief in religious evil reflected higher rates of religiosity more generally. This did not appear to be the case for African Americans. Similarly, while younger Americans also tend to hold

stronger beliefs when controlling for religion, the effect of income and education is attenuated by church attendance—with high levels of attendance neutralizing the impact of social status.

Baker's (2008) paper made a crucial contribution because it not only demonstrated that belief in religious evil could be studied empirically, but perhaps more importantly, that it *should* be studied using robust national level data. However, in providing an empirical demonstration of the importance of studying the impact of religious evil, the paper also tacitly provided a blueprint for how to study it—and that method utilized a three-item multiple measure.

Several studies have subsequently used the three-item MIM that combines belief in the Devil/Satan, demons, and hell (Ellison et al. 2021; Martinez 2013; Martinez et al. 2018). Baker et al. (2020), for example, used the three-item MIM to demonstrate a relationship between belief in religious evil and a range of attitudes associated with sexuality, including abortion, same-sex relations, pre-marital sex, extra-marital sex, pornography, and cohabitation. Elsewhere, Martinez, Tom, and Baker (2022) used the three-item measure to show that belief in religious evil is a strong predictor of support for more restrictive immigration policies, while Ellison et al. (2021) demonstrated how the MIM is a robust predictor of support for policies that expand gun rights. Jung (2020) used a two-item index (the Devil/Satan and demon items) to demonstrate that belief in religious evil is associated with higher levels of anxiety and paranoia, with Baker and Booth (2016) also adapted the index slightly by combining beliefs about the existence of Satan and hell with an item that measures agreement with whether “Satan causes most evil in the world.”

However, this emergent orthodoxy in the measurement of religious evil has recently been brought into question. Baker et al. (2020), for example, highlight that beliefs about Satan are much less nuanced than beliefs about God—where there is an abundance of research that has examined how different perceptions of God variously influence moral and political attributions (Bader et al. 2017; Froese and Bader 2010). While the historical development of the Devil is complex (Cohn 1993; Messadié 1996; Pagels 1996; Russell 1981, 1992), beliefs about the Devil/Satan are likely to be much more uniform across religious denominations *and* believers and non-believers—with the Devil being the personification of all that is in opposition to what is perceived to be good, regardless of differences in whatever good that may actually represent.

To this end, Desmond et al. (2023) have demonstrated that a single item measure of belief in Satan can be used as a robust predictor in a variety of moral domains, even when controlling for religious service attendance, biblical literalism, and images of God. This includes the “wrongness” of abortion, marital affairs, homosexuality, premarital sex, pregnancy out of wedlock, and stem cell research—with the three-item measure only significantly related in a further two arenas (cohabitation and pornography). Use of the three-item measure also did not particularly improve the adjusted  $R^2$  score, with an increase of less than 1% in 11 out of 12 domains. Similarly, the interaction effect between religious service attendance and belief in Satan was significant for 10 of the 12 moral beliefs, with higher levels of



religious attendance having little or no effect when people do not also believe in Satan.

### *The Sociology of Evil*

In addition to demonstrating the utility of a single item measure of religious evil, the results presented by Desmond et al. (2023) also suggest that where ritual is not accompanied by belief, or where belief is unsupported by ritual, then the relative influence of judgments about (deviant) morality is suppressed. This finding resonates with the more theoretical literature on what might be termed “the sociology of evil,” where several scholars have highlighted the absolute interdependence between the perception of good and the perception of evil (Alexander 2001; Douglas 2017; Lemert 1997; Wolff 1969). Alexander (2001), for example, argues that evil cannot be understood as a residual category of what is seen to be “good.” Instead, our ideas of evil—what it is, where to find it, and how to deal with it—provide the symbolic building blocks of a moral society. For Alexander, narratives of evil are the conceptual gloss on “social efforts to symbolize, narrate, code, and ritualize the good” (Alexander 2001:156). The identification and management of evil provides the contrast that is necessary to imagine how we think the world ought to be *and* an explanation for why it might not be. Therefore, the sociological analysis of evil can help to “reveal the skeletal structures upon which social communities build the stories that guide their everyday taken-for-granted political life” (Alexander 2001:166).

Similarly, Douglas (2017) notes that judgments about good and evil are categorical distinctions between fundamental absolutes. That is to say that good and evil cannot be considered to exist on a moral continuum because that is not how they are understood within everyday life. It might be possible to make an assessment about whether one evil is somehow worse than another, but neither can be considered good: “good necessarily implies a categorical contrast; if there is a good type there must be an evil type” (Douglas 2017:5). Although Douglas (2017) is much less prescriptive about the exact construction of what constitutes evil than Alexander’s (2001) consensus-based model, in both cases the contrast between an interdependent binary—good and evil—provides the framework through which moral belief and behavior can be seen to exist.

In these terms, belief in the Devil/Satan (evil) provides the necessary contrast for religious attendance (good) to be morally meaningful, and vice versa—hence the relative strength of the interaction term. The two opposing forces provide the categorical distinction for both to be meaningful.

However, while the findings of Desmond et al. (2023) appear to confirm some of the theoretical conceits presented within the sociological literature on evil, it remains to be seen whether those results can be applied in contexts outside of the USA. Indeed, while most of the work on religious evil is based on North American data, there is some evidence to warrant further investigation elsewhere. Sigelman (1977), for example, notes that the USA had higher rates of belief in the Devil and hell than other countries, while Baker and Booth (2016) also provide some

tentative evidence using the World Values Survey (2010-2014) to suggest that belief in the Devil in the USA may be twice as high as other developed countries—although they also note that these differences require further elaboration. To these ends, it is not clear whether and how the relationship between religious attendance, belief in the Devil/Satan, and morality resonates beyond the context of the USA.

### *Aims and Objectives*

This paper aims to address five key research questions. First, does the multiple item measurement of religious evil account for a greater proportion of the variance in a criterion measure of moral belief than each of its component elements? Second, do each of the components make a significant addition to the measure? Third, to what extent is the measurement of religious evil invariant across religious affiliation? Fourth, when used in conjunction with religious attendance, does the multiple item measurement of religious evil outperform single items that are designed to assess belief in the Devil/Satan? Finally, do the relationships between religious attendance, belief in the Devil/Satan, and morality resonate in contexts beyond the USA, namely the UK.

## **Methods**

### *Data*

This paper draws on data taken from CASPAR: the Chapman and Sheffield Paranormal and Religion survey (Bader, Baker, and Clark 2024). Survey respondents consisted of a random, national sample of 2,100 UK citizens, and was conducted by Ipsos Mori using their web-based panel. The panel is composed from random probability unclustered address-based sampling and has a total of over 25,000 panelists. In August, 2021 panelists were invited by email to complete the survey, with design and calibration weights then applied by the provider to correct for differences in selection probabilities and response rates between subgroups.

### *Survey Instruments*

CASPAR was designed to assess religious and alternative beliefs and experiences, while also asking respondents to respond to batteries of items on moral issues, political opinion and social issues. Items were specifically designed to resonate with measures used previously in the literature, although some adjustments were made to suit house styles and local context.

In the first instance, the analysis combines a set of indicators that measure belief in religious evil. Responding on a four-point scale (“definitely,” “probably,” “probably not,” “definitely not”), participants were asked “in your opinion, does each of the following exist?” This included—“The Devil/Satan,” “Hell,” and “Demons”—with items subsequently recoded so that a high value represents a high level of belief.

A further battery of “evil” items was adapted from a range of surveys discovered using “the measurement wizard” tool available through the Association of Religious Data Archives. This included: “Most evil in the world is caused by the Devil” (hereafter “Devil”); “Most evil in the world is caused by humankind” (“Humankind”); “Satan is the root of all evil” (“Satan\_Root”); and “Most evil in the world is caused by malignant supernatural forces” (“Supernatural forces”). These items were measured on a five-point scale: “strongly agree,” “tend to agree,” “neither agree nor disagree,” “tend to disagree,” “strongly disagree.” At the insistence of the provider, all items within the survey were accompanied by the option “prefer not to say.” All of these returns were coded as missing.

We control for a range of demographic variables within the analyses. This includes: gender (men = 1); age (16–24, 25–34, 35–44, 45–54, 55–64, 65–75); ethnicity (white = 1); marital status (married = 1); and, whether or not the respondent has children under 18 living at home (yes = 1). Education was assessed using a seven-point category system, ranging from primary school through to post-graduate levels. Income is measured on a 10-point ordinal scale (Under £5,000, £5,000–£9,999, £10,000–£14,999, £15,000–£19,999, £20,000–£24,999, £25,000–£34,999, £35,000–£44,999, £45,000–£54,999, £55,000–£99,999, £100,000 or more). England is used as the reference category for the region of the UK, with dummy variables for Wales, Scotland, and Northern Ireland.

Religious affiliation was assessed using the provider’s standard item—“What is your religion, if any?”—followed by a 17-fold classification system. To simplify the analysis, this was subsequently recoded to include: Catholic ( $n = 211$ ), Church of England ( $n = 527$ ), Protestant ( $n = 157$ ), Muslim ( $n = 88$ ), Other ( $n = 92$ ), Agnostic ( $n = 276$ ), and Atheist ( $n = 441$ ).

Religious attendance was also measured with the standard item—“How often do you attend religious services at a church, mosque, synagogue or other place of worship?”—with responses measured on an eight-point scale (never; less than once a year; once or twice a year; several times a year; once a month; two to three times a month; weekly; several times a week).

We also used eight bespoke items within the survey that were directed toward the assessment of social issues related to moral beliefs. Each of the items used the same five-point response format: “strongly agree,” “tend to agree,” “neither agree nor disagree,” “tend to disagree,” and “strongly disagree.” Within the battery, respondents were first asked “To what extent do you support or oppose each of the following things? Please select one answer per statement.” Responses were subsequently combined into four groups of two statements, with higher scores indicating support: “Congestion charges on busy roads across the UK” and “Tightening the laws in the UK to help protect the environment” (Environmental issues); “Greater funding of the National Health Service” and “Increases in the national minimum wage in the UK” (Welfare rights); “Equal rights for transgender people in the UK” and “Same-sex marriage” (Sexual freedoms); and, “The death penalty for persons convicted of murder” and “Harsher sentences than there currently are in the UK for convicted criminals” (Punitiveness).

## Analytic Strategy

Bergkvist and Rossiter (2007) suggest two principal techniques to examine the effectiveness of a multiple item measure—and both aim to establish the predictive validity of a measure through its relative relationship with a criterion variable. In respect to assessing the composition of multiple item measurements, it is possible to break down the items of a multiple-item scale and enter them as independent predictors in a regression on the criterion variable. The central item is entered first, and then each additional item in turn. If those additions provide a significant increase to the adjusted  $R^2$ , then there is evidence that the construct is suitable for multidimensional measurement (research question one). Differences between the models can be further assessed with a partial  $F$  statistic. Given that the measurement of religious evil is usually composed of three distinct components (Devil, Hell, and Demons), then it would be expected that each should make an additive contribution to the criterion measure (research question two).

Given that the three-item scale purports to measure religious evil, we might also expect that there is no appreciable difference in the performance of the scale across religious affiliation (research question three). This is sometimes referred to as measurement invariance. While there is discussion with to appropriate “cut-offs,” Item Response Theory can be used to assess both item fit and Differential Item Functioning (DIF) (see Engelhard and Wang 2021). Item fit is an assessment of the suitability of an item for use in the model. The most robust assessments of item fit using a Rasch model require infit and outfit statistics of between .7 and 1.3 (Aryadoust, Ng, and Sayama 2021), with violations of infit thought to be more problematic than outfit violations (Linacre 2002). On the other hand, Differential Item Functioning occurs when an item within a latent trait varies systematically between groups. Following Pietryka and MacIntosh (2022), it is possible to detect DIF by examining group differences in the standardized residuals using a one-way ANOVA test—with a significant result suggesting the presence of DIF on a particular item. Item contrasts can then be used to identify differences between individual groups.

There are several methods that can be used to establish the utility of alternative SIMs against their multiple item counterpart. Perhaps the most common is to assess the convergent validity that exists between single and multiple item versions. Unfortunately, there is little agreement on what value represents acceptable convergence. Greiff and Allen (2018) suggest that one approach is to adopt alpha coefficients associated with test-retest reliability, where  $r \geq .90$  is excellent,  $\geq .80$  is good, and  $\geq .70$  is moderate. However, these cut-offs are only general rules of thumb.

Following Gorsuch and McFarland (1972), it should also be possible to inspect the commonalities and eigenvalues associated with the SIMs and MIMs after performing a principal components analysis. This would give an indication of the extent to which each individual scale had variance in common with the other scales in the analysis. Again, while cutoffs are general rules of thumb, it could be expected that each of the scales would load on to a single factor that explained at least 70%

of the variance, with commonalities of over .60 for each item (MacCallum et al. 2001), and component loadings of over .71 (Comrey and Lee 2013).

However, both alpha and PCA will only confirm the relative similarity of the items. For Bergkvist and Rossiter (2007), the relative influence on a criterion measure is the most effective method of establishing predictive validity. That is to say that if a SIM is comparable in performance with a MIM, then there is a strong case for the single item approach to be retained. To this end, it would be possible to accept the single item measure by making three further assessments. Firstly, that there is no appreciable decline in the adjusted  $R^2$  score when using the single item measure. Secondly, there is little difference in the composition of the model when controlling for demographic characteristics. Third, given that both the theoretical discussion and empirical evidence suggest an interdependence between good and evil, we would also expect an interaction between religious evil and religious attendance (research question four). These final outputs should be sufficient to examine the usefulness of a single item measure, and the extent to which the relationships between religious attendance, belief in the Devil/Satan, and moral belief exist in the context of the UK (research question five).

## Results

The results of the OLS hierarchical regression examining the relative impact of each item of the multiple item measure of religious evil are depicted in Tables 1 and 2. Four moral domains are assessed: welfare support and sexual freedoms feature in Table 1, and environmental issues and punitiveness in Table 2.<sup>1</sup>

The results presented in Table 1 (columns 1–4) suggest that only the item assessing belief in the Devil/Satan provides a positive and significant increase beyond the controls in support for welfare issues. Columns 5 to 8, however, suggest that while belief in the Devil/Satan is negatively associated with sexual freedoms (support of trans/gay rights), the effect is negated by the inclusion of other variables. Belief in Demons is the only evil measure that remains significant in the final model. The results presented in the first four columns of Table 2 suggest that none of the items on religious evil provide a significant increase to  $R^2$ —although it is, perhaps, notable that the Hell variable makes a (non-significant) negative contribution to the final model.

While belief in the Devil/Satan is an improvement on the null model, only belief in Demons makes a significant contribution to a final model predicting greater support for harsher punitive policy (columns five to eight). Across all the final models, no significant increase in adjusted  $R^2$  is reported for the item measuring belief in Hell, while the performance of the Devil/Satan and Demons items is variable. Therefore, within the context of the UK, there is some evidence to suggest that these items might not be suitable for multidimensional measurement (research question one), because they can act in different ways when regressed on criterion measures (research question two).

To assess measurement variance using Item Response Theory, we used the TAM package in R (Robitzsch, Kiefer, and Wu 2024) and followed the method for

**Table 1.** Regression of Multiple Item Measures of Religious Evil on Moral Beliefs (1).

Variables	Welfare issues				Sexual freedoms			
	Model 1	Model 2	Model 3	Model 4	Model 1	Model 2	Model 3	Model 4
(Intercept)	8.77***	8.54***	8.54***	8.58***	8.12***	8.69***	8.81***	8.91***
Gender	-.49***	-.48***	-.48***	-.48***	-.61***	-.63***	-.64***	-.64***
Ethnicity	.48**	.50**	.50**	.50**	.45*	.41*	.41*	.41*
Age	.00	.01	.01	.01	-.16***	-.19***	-.20***	-.20***
Education	-.00	.00	.00	-.00	.12***	.12***	.12***	.12***
Income	-.05*	-.04*	-.04*	-.04*	.09***	.08***	.08***	.08***
Married	-.12	-.13	-.13	-.14	-.30**	-.28**	-.28**	-.29**
Children	-.04	-.05	-.05	-.04	-.18	-.17	-.15	-.13
Wales	.03	.04	.04	.05	.12	.09	.07	.08
Scotland	.35*	.36*	.36*	.35*	.23	.20	.21	.17
NI	-.15	-.15	-.15	-.13	-.34	-.34	-.32	-.28
Catholic	-.11	-.25	-.25	-.24	-1.07***	-.76***	-.67***	-.64***
CoE	-.22	-.31*	-.31*	-.30*	-.76***	-.54***	-.47***	-.46***
Protestant	-.38*	-.54**	-.54**	-.53**	-1.65***	-1.26***	-1.21***	-1.19***
Muslim	-.57*	-.75**	-.75**	-.75**	-2.12***	-1.71***	-1.55***	-1.55***
Other	.10	.01	.01	.04	-.50*	-.28	-.27	-.19
Agnostic	.01	-.02	-.02	-.01	-.26	-.20	-.19	-.18
Satan		.10*	.10	.17*		-.25***	-.08	.07
Hell			-.00	.03			-.23**	-.16
Demons				-.11				-.27***
<i>n</i>	1,532	1,532	1,532	1,532	1,522	1,522	1,522	1,522
<i>R</i> <sup>2</sup> / <i>R</i> <sup>2</sup> adjusted	.06/.05	.07/.05	.07/.05	.07/.06	.18/.17	.19/.18	.20/.19	.20/.19
Partial <i>F</i>	n/a	5.26*	0.00	2.60	n/a	24.05***	8.78***	11.31***
		( <i>df</i> = 1;1,514)	( <i>df</i> = 1; 1,513)	( <i>df</i> = 1; 1,512)		( <i>df</i> = 1;1,504)	( <i>df</i> = 1;1,503)	( <i>df</i> = 1;1,502)

\**p* < .05. \*\**p* < .01. \*\*\**p* < .001.

**Table 2.** Regression of Multiple Item Measures of Religious Evil on Moral Beliefs (2).

Variables	Environmental issues				Punitiveness			
	Model 1	Model 2	Model 3	Model 4	Model 1	Model 2	Model 3	Model 4
(Intercept)	6.88***	6.68***	6.69***	6.67***	7.50***	7.08***	6.99***	6.92***
Gender (Male)	.17	.18	.18	0.17	.03	.05	.06	.05
Ethnicity	.29	.18	.19	0.18	.52*	.28	.22	.20
Age	-.05	-.13	-.12	-0.12	.84***	.68***	.64***	0.62***
Education	-.16	-.30	-.29	-0.29	-.00	-.29	-.33	-.34
Income	-.63*	-.78**	-.77**	-0.77**	.56	.25	.15	.14
Married	.33	.26	.26	0.24	.77**	.60*	.59*	.53
Children	-.16	-.18	-.18	-0.18	-.05	-.10	-.10	-.11
Wales	-.24*	-.24*	-.24*	-0.24*	.09	.08	.07	.08
Scotland	-.23*	-.22*	-.22*	-0.23*	.03	.05	.06	.05
NI	.00	.00	.00	0.00	.03	.04	.04	.04
Catholic	.10**	.11***	.11***	0.11***	-.30***	-.29***	-.30***	-.29***
CoE	-.04	-.03	-.03	-0.03	.11*	.13**	.13**	.13**
Protestant	-.02	-.02	-.02	-0.03	.21	.20	.19	.18
Muslim	-.35	-.33	-.34	-0.34	-.06	-.04	-.03	-.03
Other	-.07	-.07	-.07	-0.06	-.19	-.17	-.17	-.15
Agnostic	.10	.10	.10	0.09	-.40	-.40	-.41	-.44
Satan		.09	.11	0.06		.19**	.07	-.05
Hell			-.02	-0.04			.16	.10
Demons				0.08				0.20*
<i>n</i>	1,528	1,528	1,528	1,528	1,525	1,525	1,525	1,525
<i>R</i> <sup>2</sup> / <i>R</i> <sup>2</sup> adjusted	.03/.02	.04/.02	.04/.02	.04/.02	.09/.08	.10/.09	.10/.09	.10/.09
Partial <i>F</i>	n/a	3.03 ( <i>df</i> = 1; 1,510)	0.89 ( <i>df</i> = 1; 1,509)	0.94 ( <i>df</i> = 1; 1,508)	n/a	9.07** ( <i>df</i> = 1; 1,507)	3.01 ( <i>df</i> = 1; 1,506)	4.48* ( <i>df</i> = 1; 1,505)

\**p* < .05. \*\**p* < .01. \*\*\**p* < .001.

**Table 3.** Religious Evil: Item Fit Statistics.

Item	Outfit	Outfit_t	Infit	Infit_t
Satan	0.661	-5.608***	0.769	-6.966***
Hell	0.797	-3.331***	0.890	-3.171**
Demons	0.779	-3.498***	0.881	-3.496***

\*\* $p < .01$ . \*\*\* $p < .001$ .

**Table 4.** Religious Evil: Differential Item Functioning by Religious Affiliation.

Item	Df	Sumsq	Meansq	Statistic
Satan	6	42.849	7.141	22.784***
Hell	6	55.337	9.223	23.765***
Demons	6	13.851	2.308	5.536***

\*\*\* $p < .001$ .

polytomous data outlined by Pietryka and MacIntosh (2022). Item fit statistics for the variables within the religious evil MIM are presented in Table 3. As previously suggested, the most robust assessments require infit and outfit statistics of between .7 and 1.3. Further, there is a general acceptance within the literature that large sample sizes ( $>n = 500$ ) will tend to bias t-statistics for polytomous data, whereas mean square statistics have been shown to be relatively stable (Smith et al. 2008). Given that the sample size in this instance is large, we discount the significance levels associated with outfit and infit measures. The outfit score for the Satan variable does not meet the robust cut-off, but this is marginal and unlikely to be detrimental in “low stakes” measurements. All the items score below 1 and therefore fit the model better than might otherwise be expected.

However, inspection of Table 4 reveals evidence of Differential Item Functioning (DIF) for all the items in relation to Religious Affiliation. To be clear, there is some recognition in the literature that DIF is context dependent, and there is difference between benign and adverse DIF. Benign DIF is generally not considered to be detrimental to the scale, while adverse DIF is typically attributed to measurement error. As suggested by Gierl et al. (2001:167): “When conducting DIF analyses, it is a matter of judgment as to whether the secondary dimension is interpreted as benign or adverse in a particular testing situation ... the purpose of the test, the nature of the secondary dimension, and the examinees of interest must be considered.”

In this case, it is possible to make at least two clear points in relation to the argument that the Differential Item Functioning is adverse. Examination of Table 5 suggests that there are no significant differences in item functioning on the Demon item between those claiming Catholic, Protestant, and CoE affiliation and those claiming to be agnostic. This is not the case for either beliefs in Satan or Hell where



**Table 5.** Item Contrasts by Religious Affiliation.

Affiliation #1	Affiliation #2	Satan	Hell	Demons
CoE	Catholic	-0.126	-0.067	-0.022
Protestant	Catholic	0.158	-0.012	-0.004
Muslim	Catholic	0.162	0.379***	-0.024
Other	Catholic	-0.069	-0.254*	0.166
Atheist	Catholic	-0.320***	-0.314***	-0.191*
Agnostic	Catholic	-0.259***	-0.327***	-0.082
Protestant	CoE	0.285***	0.056	0.017
Muslim	CoE	0.288***	0.446***	-0.002
Other	CoE	0.057	-0.187	0.188
Atheist	CoE	-0.194***	-0.246***	-0.170***
Agnostic	CoE	-0.133*	-0.260***	-0.060
Muslim	Protestant	0.004	0.391***	-0.019
Other	Protestant	-0.228*	-0.242	0.170
Atheist	Protestant	-0.478***	-0.302***	-0.187*
Agnostic	Protestant	-0.418***	-0.315***	-0.078
Other	Muslim	-0.231	-0.633***	0.190
Atheist	Muslim	-0.482***	-0.693***	-0.168
Agnostic	Muslim	-0.421***	-0.706***	-0.058
Atheist	Other	-0.251**	-0.060	-0.357***
Agnostic	Other	-0.190	-0.073	-0.248*
Agnostic	Atheist	0.061	-0.013	0.109

\* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$ .

significant differences are reported, and perhaps would be expected. This means that where latent trait scores are the same for individuals, there are no systematic differences across the groups on the Demon item. This would seem to compromise the argument that Demons item is a distinguishing feature of a specifically *religious* evil.

Perhaps more problematically, Table 5 also suggests evidence of a consistent difference between Muslims and other forms of affiliation in respect to the Hell item. Again, this is to suggest that where individuals in these groups correspond on the scale overall, they will likely differ in respect to their belief in Hell—with Muslims more likely to score higher on the item.

As a result, there is evidence to suggest that the items within the religious evil MIM are not invariant across religious affiliations (research question three). That is to say that belief in the Devil/Satan, belief in Hell and belief in Demons operate in different ways across religious groups.

To assess the convergent reliability of the single item measures, Table 6 provides a correlation matrix between all individual measures and the multiple item counterpart. All correlations are statistically significant ( $p < .01$ ). The alpha coefficient between these items is .681 ( $n = 1768$ ) and can be considered moderate. However, it is notable that the correlations between the “Most evil in the world is caused by humankind” (Humankind) variable and every other measure of religious evil are

**Table 6.** Correlation Matrix of Evil Items.

Item	MIM	Satan_Root	Devil	Supernatural forces	Humankind
MIM	1	.595***	.672***	.599***	-.118***
Satan_Root		1	.632***	.575***	-.191***
Devil			1	.727***	-.241***
Supernatural Forces				1	-.224***
Humankind					1

\*\*\* $p < .001$ .

**Table 7.** PCA Component Loadings.

Item	Component loading
MIM	0.841
Satan_Root	0.818
Devil	0.891
Supernatural forces	0.852

negative. If the Humankind variable is removed, the alpha coefficient also rises to .794. These results suggest that the religious evil index, “Satan is the root of all evil” (Satan\_Root), “Most evil in the world is caused by the Devil” (Most evil ... Devil), and “Most evil in the world is caused by malignant supernatural forces” (Supernatural Forces) items are similar in nature, but different to the “Humankind” measure.

Principal components analysis also reveals communalities over .60 for each item, except the “Humankind” item, which achieved .11 and was therefore removed from the analysis. With an eigenvalue of 2.90, the results of the PCA suggest a single factor solution that explains 72.4% of the variance, with each component loading above .80 (see Table 7). So, while the MIM, “Most evil ... Devil,” “Satan-Root,” and “Supernatural forces” items appear suitable for comparison, the item assessing “Humankind” may operate in a different way from the other items as it does not load on to the single factor identified in the solution. However, we include it in the remaining analyses to continue to assess its performance in relation to moral beliefs.

Tables 8 to 11 show the results of the OLS hierarchical regressions examining the relative performance of each measure of each evil item in combination with religious attendance. Again, regression coefficients are used to assess the relative impact of the independent measures on the criterion measures—environmental issues, welfare issues, sexual freedoms, and punitiveness.<sup>2</sup>

In respect to environmental issues, Table 8 demonstrates that while none of the items measuring religious evil achieve significance individually, when used in combination with religious attendance, the interactive terms for the “Most evil in the world is caused by the Devil” item is the only measure to make significant contributions to the model—with the final model explaining 6% more of the variance in  $R^2$  adjusted than the MIM.

**Table 8.** Comparative Performance of Items by Support for Environmental Issues.

Variables	MIM	Belief_Satan	Satan_Root	Devil	Supernatural forces	Humankind
(Intercept)	6.85***	6.80***	6.79***	6.99***	6.82***	6.62***
Gender	-0.21*	-0.21*	-0.21*	-0.24**	-0.24*	-0.21*
Ethnicity	0.19	0.19	0.20	0.25	0.24	0.19
Age	-0.03	-0.03	-0.04	-0.04	-0.03	-0.06
Education	0.11***	0.11***	0.11***	0.11***	0.11***	0.10**
Income	0.01	0.01	0.01	0.01	0.01	0.00
Married	-0.27**	-0.27**	-0.27**	-0.27**	-0.26**	-0.25*
Children	-0.00	0.01	-0.01	0.01	-0.01	0.01
Wales	-0.23	-0.23	-0.25	-0.25	-0.25	-0.21
Scotland	-0.05	-0.06	-0.07	-0.05	-0.04	-0.07
NI	0.11	0.14	0.10	0.15	0.12	0.08
Catholic	0.19	0.19	0.20	0.29	0.23	0.26
CoE	-0.08	-0.09	-0.09	0.01	-0.06	-0.04
Protestant	-0.23	-0.25	-0.18	-0.12	-0.16	-0.08
Muslim	-0.68*	-0.67*	-0.64*	-0.48	-0.55	-0.48
Other	0.26	0.26	0.27	0.34	0.27	0.27
Agnostic	-0.17	-0.18	-0.17	-0.15	-0.17	-0.17
Religious attendance	-0.10	-0.09	-0.08	-0.16**	-0.10	-0.08
Evil item	0.00	0.02	0.01	-0.11	-0.04	0.06
Evil item by Rel. Att.	0.01	0.03	0.02	0.05***	0.04*	0.02
Observations	1,477	1,477	1,477	1,477	1,477	1,477
R <sup>2</sup> /R <sup>2</sup> adjusted	.035/.022	.035/.022	.034/.022	.040/.028	.036/.023	.036/.024

\* $p < .05$ , \*\* $p < .01$ , \*\*\* $p < .001$ .

Therefore, the “Most evil ... Devil” variable may be more sensitive than other measures when combined with religious attendance.

Relative performance with respect to support for welfare is assessed in Table 9. Beyond the controls, the “Satan is the root of all evil,” “Most evil ... Devil” and “Most evil ... Humankind” items achieve significance in themselves. However, whereas the “Most evil ... Devil” and “Satan ... root” items are negatively associated with welfare support, the “Humankind” item is positive. This again provides some further evidence that these items are indicative of substantively different moral structures. Only the interaction terms for the MIM, the single item of belief in the Devil/Satan, and the “Most evil ... Devil” variable achieve significance. However, the MIM explains 4% more of the variance in  $R^2$  adjusted than the “Most evil ... Devil” variable, and the single-item measure of belief 10% more. So, while both the single-item measure for belief in the Devil/Satan and the “Most evil ... Devil” variable are just as sensitive as the multiple-item measure—each determining that people with lower religious attendance who don’t believe in the Devil/Satan are more likely to support increased welfare—the single item explains slightly more of the variance.

Table 10 shows the comparative results with respect to attitudes toward sexual freedoms. Again, the “Humankind” item is the only measure to demonstrate a positive association with the criterion measure, with all other measures of religious evil being negative—although only the “Satan ... Root,” “Most evil ... Devil,” and “Supernatural forces” variables make significant additions to their respective models. However, only the “Most evil ... Devil” and “Supernatural forces” items have significant interactions with religious attendance, explaining a total of 23% and 22% of the variance in adjusted  $R^2$  respectively. Therefore, the “Most evil ... Devil” item is more sensitive than both the multiple-item measure and the single belief item measure in determining that people with less religious attendance who do not believe in the Devil are more likely to support gay/trans rights.

Finally, Table 11 shows the comparative results in performance on the punitive criterion measure. The “Humankind” item is the only measure that does not demonstrate a significant, positive association with the criterion measure, whereas the model containing “Supernatural forces” is the only model where religious attendance achieves significance. None of the interaction measures are significant. The total variance explained is highest for the “Supernatural forces” and “Satan\_Root” items (11.0% and 10.8% respectively), while the multiple item measure (10.1%) marginally outperforms both the single item belief in the Devil/Satan (9.5%) measure and the “Most evil ... Devil” item (9.7%). Religious attendance, therefore, does not appear to predict support for harsher punitive policy, although belief in the Devil is a significant predictor in its own right.

## Discussion

There is an emerging body of literature across several disciplines to suggest that single-item measures can be just as effective as their multiple-item counterparts. At the same time, there is also some emerging evidence to suggest that multiple-item

**Table 9.** Comparative Performance of Items by Support for Welfare.

Variables	MIM	Belief_Satan	Satan_Root	Devil	Supernatural forces	Humankind
(Intercept)	8.95***	8.95***	9.15***	9.32***	9.04***	8.10***
Gender	-.49***	-.48***	-0.51***	-0.52***	-0.50***	-0.48***
Ethnicity	0.47**	0.47**	0.47**	0.44**	0.45**	0.46**
Age	0.01	0.01	-0.01	-0.01	-0.01	-0.03
Education	0.01	0.01	-0.00	-0.00	-0.00	-0.00
Income	-0.04*	-0.04*	-0.05*	-0.05*	-0.05*	-0.05*
Married	-0.13	-0.14	-0.11	-0.11	-0.11	-0.09
Children	0.01	0.02	0.03	0.04	0.03	0.04
Wales	0.05	0.05	0.03	0.03	0.05	0.08
Scotland	0.33*	0.32*	0.34*	0.31*	0.31*	0.31*
NI	-0.10	-0.06	-0.10	-0.10	-0.11	-0.17
Catholic	0.03	0.02	0.20	0.25	0.14	0.13
CoE	-0.10	-0.10	0.00	0.02	-0.06	-0.03
Protestant	-0.32	-0.36	-0.12	-0.07	-0.17	-0.09
Muslim	-0.31	-0.32	-0.03	-0.02	-0.13	-0.02
Other	0.09	0.09	0.17	0.19	0.15	0.10
Agnostic	0.03	0.03	0.07	0.07	0.04	0.02
Religious attendance	-0.25***	-0.26***	-0.10	-0.16***	-0.06	-0.15*
Evil item	-0.01	-0.02	-0.11*	-0.20**	-0.06	0.21**
Evil item by Rel. Att.	0.02**	0.06***	0.01	0.03*	-0.00	0.02
Observations	1,482	1,482	1,482	1,482	1,482	1,482
R <sup>2</sup> /R <sup>2</sup> adjusted	.077/.065	.083/.071	.071/.059	.073/.061	.069/.057	.094/.083

\* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$ .

**Table 10.** Comparative Performance of Items by Support for Sexual Freedoms.

Variables	MIM	Belief_Satan	Satan_Root	Devil	Supernatural forces	Humankind
(Intercept)	8.71***	8.63***	9.16***	9.81***	9.63***	7.44***
Gender	-0.55***	-0.55***	-0.58***	-0.60***	-0.59***	-0.52***
Ethnicity	0.32	0.32	0.32	0.18	0.26	00.33
Age	-0.20***	-0.19***	-0.20***	-0.22***	-0.21***	-0.20***
Education	0.12***	0.13***	0.11***	0.10**	0.11***	0.12***
Income	0.08***	0.08***	0.08***	0.08***	0.08***	0.09***
Married	-0.24*	-0.24*	-0.24*	-0.24*	-0.26*	-0.24*
Children	-0.14	-0.15	-0.12	-0.07	-0.10	-0.13
Wales	0.15	0.17	0.16	0.17	0.23	0.22
Scotland	0.25	0.26	0.31	0.22	0.21	0.28
NI	-0.23	-0.26	-0.22	-0.27	-0.25	-0.28
Catholic	-0.46*	-0.52**	-0.35	-0.16	-0.31	-0.58**
CoE	-0.36*	-0.40**	-0.22	-0.14	-0.20	-0.39**
Protestant	-0.89***	-0.94***	-0.86***	-0.62**	-0.75***	-0.99***
Muslim	-1.19***	-1.30***	-1.11***	-1.10***	-1.26***	-1.36***
Other	-0.04	-0.08	0.00	0.07	0.08	-0.15
Agnostic	-0.17	-0.18	-0.13	-0.11	-0.14	-0.22
Religious attendance	-0.03	-0.08	-0.15*	-0.31***	-0.32***	-0.14
Evil item	-0.05	-0.09	-0.27***	-0.57***	-0.51***	0.26***
Evil item by Rel. Att.	-0.01	-0.03	0.00	0.07***	0.07***	-0.01
Observations	1,471	1,471	1,471	1,471	1,471	1,471
R <sup>2</sup> /R <sup>2</sup> adjusted	.214/.203	.208/.198	.225/.215	.240/.230	.231/.221	.216/.206

\* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$ .

**Table 11.** Comparative Performance of Items by Support for Harsher Punitive Policy.

Variables	MIM	Belief_Satan	Satan_Root	Devil	Supernatural forces	Humankind
(Intercept)	6.84***	6.99***	6.84***	6.93***	6.76***	7.60***
Gender	0.07	0.06	0.08	0.06	0.04	0.04
Ethnicity	0.06	0.06	0.06	0.15	0.18	0.05
Age	0.14***	0.13**	0.14**	0.13**	0.15***	0.10*
Education	-0.29***	-0.29***	-0.27***	-0.28***	-0.27***	-0.30***
Income	0.04	0.04	0.04	0.04	0.04	0.03
Married	0.09	0.09	0.08	0.08	0.09	0.11
Children	0.24	0.26	0.22	0.22	0.20	0.26
Wales	0.01	-0.00	-0.01	-0.02	-0.08	-0.01
Scotland	-0.19	-0.21	-0.27	-0.19	-0.14	-0.24
NI	-0.35	-0.34	-0.36	-0.30	-0.32	-0.37
Catholic	0.37	0.45	0.39	0.47*	0.45*	0.71**
CoE	0.71***	0.76***	0.68***	0.80***	0.77***	0.96***
Protestant	-0.09	-0.03	-0.00	0.02	-0.00	0.32
Muslim	0.53	0.64	0.54	0.80*	0.84*	1.04**
Other	0.53	0.58	0.60*	0.67*	0.58	0.76*
Agnostic	-0.13	-0.11	-0.14	-0.10	-0.10	-0.05
Religious attendance	-0.10	-0.07	-0.13	-0.08	-0.16*	-0.19
Evil item	0.12***	0.29**	0.29***	0.26**	0.30***	0.00
Evil item by Rel. Att.	-0.00	-0.02	-0.00	-0.01	0.01	0.03
Observations	1,481	1,481	1,481	1,481	1,481	1,481
R <sup>2</sup> /R <sup>2</sup> adjusted	.112/.101	.107/.095	.120/.108	.108/.097	.121/.110	.100/.088

\* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$ .

measures of religious evil can effectively be reduced to a single item—belief in the Devil—without significant detrimental impact on explanatory capacity. This paper makes an important contribution to this literature by providing further evidence to demonstrate that single-item measurements of religious evil can be just as effective as their multiple-item counterparts. While there is some variation in performance across items and between criterion variables, the multiple-item measure for religious evil is neither consistently better in terms of the amount of variance explained nor in terms of achieving significance than other measures. Differences in performance between the item “Most evil in the world is caused by the Devil” and both the single-item measure of belief in the Devil and the multiple-item measure of religious evil are marginal, particularly when used in conjunction with religious attendance. Moreover, there is evidence that the religious evil MIM should not be assumed to be invariant across religious affiliation and may conflate nuanced differences in response. Therefore, these findings offer further support to the idea that belief in the Devil should be considered a “doubly concrete” concept that may be more usefully assessed with a single item. In addition, there is also some evidence offered here that the “Most evil is caused by humankind” item may interrelate with moral beliefs in ways different to beliefs in the Devil. Further exploration of other measures assessing belief in evil may be advantageous.

There are, of course, limitations to the findings. The first concerns the lack of wider cross-cultural evidence pertaining to items assessing religious evil. Alongside Desmond et al. (2023), this paper demonstrates that there is evidence of effectiveness of single items in the measurement of religious evil in the UK and the USA. However, while there are certainly differences between the two countries, they do have somewhat intertwined histories and it remains to be seen how items assessing belief in the devil, religious evil, religious attendance, and morality interact within different, wider contexts. Further cross-cultural evidence would be beneficial to help specify the utility of single and multiple-item measures, their relative interaction with religious affiliation and attendance, and how they work in conjunction with social and moral beliefs. It is also not possible to eliminate the possibility of a method effects within the measurement of these items. Method effects occur where variation in an item can be explained due to the method of measurement, rather than being a property of the attribute (Schweizer 2020). The measures in this study were delivered in two distinct batteries: the initial belief items (Devil, Hell, Demons), and a battery focused on the measurement of belief in evil. Evidently, this paper does report fine-grained distinctions between the performance of these items and this is strongly indicative that there are differences between their relative performance. However, while a general assumption is made within survey research that each item is independent of another, further research using these items in different contexts would help to discount the impact of potential methods effects. Moreover, when used in conjunction with “evil” items, it also remains to be seen whether other items that symbolize the “good”—whether religious belief, practice, or experience—are more sensitive than religious attendance. Indeed, there is some well-established discussion concerning the limitations of religious attendance as an



indicator of religiosity and it may be that there are more sensitive measures that can be used in conjunction with measures of religious evil (Smith 1998).

Regardless of these limitations, the paper makes four important contributions to the literature. In the first instance, there are some trends apparent in the United Kingdom that seem to be congruent with those in North America. If we were to take the “Devil” item as a barometer of the relationship between religious evil and moral belief, there is evidence here to suggest that increased support in respect to environmental issues is related to decreases in religious attendance, particularly when accompanied by little or no belief in the Devil. This appears to also be the case for support for increased welfare support and attitudes toward sexual freedoms; religious attendance and belief that “the Devil is responsible for most evil in the world” suppress support for each, particularly when they occur together. In contrast, support for harsher punitive policy is only associated with higher levels of belief in the Devil. While this finding might appear somewhat different to other criterion measures, the nuanced relationship between religious evil and punitiveness has been reported elsewhere (Baker and Booth 2016). These findings provide support for further investigation of this relationship.

In demonstrating the interaction between belief in the Devil and religious attendance, the paper also provides further support for the absolute interdependence between ideas of good and evil. Of course, it is possible to attend religious services and not believe in the Devil. Similarly, it is possible to believe in the Devil and be largely irreligious. However, it is evident that the presence of one supports the meaning and interpretation of the other. Attendance demonstrates a ritual “good,” but that alone does not necessarily have the contrast to propel understandings of those actions into a meaningful moral sphere. Resonating with some long-standing, but largely untested, theoretical conceits in the sociology of evil (Alexander 2001; Douglas 2017), belief in the Devil appears to be the catalyst that, when combined with a ritual “good” in the form of religious attendance, allows moral attributions to be made with greater certainty. That is to say that if a person believes that they are good, and that there is an evil agonist in the world, then there is a greater platform to interpret morality in a binary format. Evidently, belief in greater punitiveness is something of an outlier in this respect where belief in the Devil is sufficient for support. However, and as previously stated, this has not gone unnoticed elsewhere and focused research is likely to be necessary to disentangle these effects. What this paper does do, however, is further demonstrate the importance of evil in understanding punitive ideology beyond the context of the USA.

Third, while the results in this paper suggest that the single-item use of the Devil variable may be beneficial, this does not mean that the items assessing belief in Hell or Demons are unimportant. These variables may also be “doubly concrete” in as much as they assess very specific beliefs that may also vary by religious affiliation. Such nuances are also likely to be underestimated within approaches that package belief as a general index. Further examination of how these items might vary across affiliation is likely to be particularly valuable.

Finally, this paper offers evidence to suggest that further empirical assessment concerning more secular beliefs in evil is necessary. This includes the relative impact

on moral beliefs, but also the different forms these beliefs might take. The differential performance of the “most evil in the world is caused by humankind” suggests the possibility that belief in a distinctly human evil is separate from belief in a religious evil, however it is measured. In the context of a secular society where evil is often associated with serial killers, child murderers, terrorists, and pedophiles, this might not be too surprising. It has, however, gone largely unexplored within the literature. Of course, Augustinian metaphysics attempted to resolve the Epicurean paradox by placing the cause of evil within humankind and it would not necessarily be contradictory to believe that “the Devil is the cause of most evil in the world” and “most evil in the world is caused by humankind” (Evans 1982). That said, and unlike the “Most evil ... Devil” item, the evidence in this paper suggests that those who believe in human evil are more likely to support welfare issues and greater sexual freedoms, and although there is a non-significant relationship between belief in human evil and environmental issues, it is notable that the relationship is a positive one. To put this more simply, the belief that most evil in the world is caused by the Devil appears to lead to different outcomes than the belief that “most evil is caused by humankind.” Further investigation is, therefore, necessary to explore the nature, form, and influence of more secular beliefs in evil. Indeed, the results presented within this paper again highlight the importance of further developing robust survey items relating to beliefs of both “good” and “evil” across cross-cultural contexts and their relative influence on moral beliefs.

### **Declaration of Conflicting Interests**

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.


### **Funding**

The author(s) received no financial support for the research, authorship, and/or publication of this article.

### **Ethical Approval**

The project received ethical approval from the Chapman University Institutional Review Board (IRB-21-275) and written informed consent was given to the survey provider by participants in all cases.

### **ORCID iD**

Tom Clark  <https://orcid.org/0000-0001-6871-629X>

### **Data Availability Statement**

The data in this study are available upon reasonable request.

## Notes

1. All variables satisfied conventional cut-offs for VIF ( $<5$ ).
2. To assess whether the relative performance of the variables associated with evil were consistent across the boundaries of Christianity, we also performed the analysis contained in Tables 8 to 11 using an interaction model that allowed the evil items to interact with a “Christianity” dummy variable. No significant effects were observed for either the dummy variable or the moderating variable in respect to either support for greater welfare or sexual freedoms. However, while there was no significant effect for the (negative) Christianity dummy in respect to environmental issues, there was a positive interaction effect for the Christian dummy and the evil belief item in respect to the MIM ( $p < .05$ ), belief in Satan/Devil ( $p < .05$ ), Satan\_Root ( $p < .05$ ), and Supernatural Forces ( $p < .01$ ) variables. No interaction effect is reported for the “Most evil ... Devil” variable. Being affiliated with Christianity was also positively related with increased support for punitiveness, with the MIM ( $p < .001$ ), belief in the Devil/Satan ( $p < .001$ ), Satan\_Root ( $p < .001$ ), Devil ( $p < .001$ ), and Supernatural Forces ( $p < .01$ ) variables. Significant (negative) interaction effects are also reported for the MIM ( $p < .001$ ), belief in the Devil/Satan ( $p < .001$ ), Satan\_Root ( $p < .01$ ), and Devil ( $p < .01$ ) variables, with a positive interaction effect reported for the Humankind variable ( $p < .01$ ). These results tend to confirm the nuanced nature of the relationship that exists between ideas of evil, religious affiliation, and punitiveness, while also providing some further support for the use of “the Most evil in the world is caused by the Devil” variable across religious contexts.

## References

- Alexander, Jeffrey C. 2001. “Towards a Sociology of Evil.” Pp. 153–72 in *Rethinking Evil: Contemporary Perspectives*, edited by M. P. Lara. Los Angeles, CA: University of California Press.
- Allen, Mark S., Dragos Iliescu, and Samuel Greiff. 2022. “Single Item Measures in Psychological Science.” *European Journal of Psychological Assessment* 38(1):1–5. doi:10.1027/1015-5759/a000699.
- Aryadoust, Vahid, Li Y. Ng, and Hiroki Sayama. 2021. “A Comprehensive Review of Rasch Measurement in Language Assessment: Recommendations and Guidelines for Research.” *Language Testing* 38(1):6–40. doi:10.1177/0265532220927487.
- Bader, Christopher D., Joseph O. Baker, and Tom Clark. 2024. “CASPAR: The Chapman and Sheffield Paranormal and Religion Survey (2021).”
- Bader, Christopher D., Paul Froese, Byron Johnson, F. Carson Mencken, and Rodney Stark. 2023. “Baylor Religion Survey, Wave I (2005).” *Data Archive*. doi:10.17605/OSF.IO/B9H7X.
- Bader, Christopher D., F. Carson Mencken, and Paul Froese. 2007. “American Piety 2005: Content and Methods of the Baylor Religion Survey.” *Journal for Scientific Study of Religion* 46(4):447–63. doi:10.1111/j.1468-5906.2007.00371.x.
- Bader, Christopher D., Andrea Molle, Benjamin Gurrentz, and James J. Coyle. 2017. “For the Wrath of God: Fatalism and Images of God in Violent Regions of the World.” *Sociology of Religion* 78(2):173–91. doi:10.1093/socrel/srx010.
- Baker, Joseph O. 2008. “Who Believes in Religious Evil? An Investigation of Sociological Patterns of Belief in Satan, Hell, and Demons.” *Review of Religious Research* 50(2): 206–20.

- Baker, Joseph O. and Alexis L. Booth. 2016. "Hell to Pay: Religion and Punitive Ideology among the American Public." *Punishment & Society* 18(2):151–76. doi:10.1177/1462474516635886.
- Baker, Joseph O., Andrea Molle, and Christopher D. Bader. 2020. "The Flesh and the Devil: Belief in Religious Evil and Views of Sexual Morality." *Review of Religious Research* 62(1):133–51. doi:10.1007/s13644-020-00403-4.
- Bergkvist, Lars. 2015. "Appropriate Use of Single-Item Measures is Here to Stay." *Marketing Letters* 26(3):245–55. doi:10.1007/s11002-014-9325-y.
- Bergkvist, Lars. 2016. "The Nature of Doubly Concrete Constructs and How to Identify Them." *Journal of Business Research* 69(9):3427–29. doi:10.1016/j.jbusres.2016.02.001.
- Bergkvist, Lars and John R. Rossiter. 2007. "The Predictive Validity of Multiple-Item Versus Single-Item Measures of the Same Constructs." *Journal of Marketing Research* 44(2):175–84. doi:10.1509/jmkr.44.2.175.
- Bergkvist, Lars and John R. Rossiter. 2009. "Tailor-Made Single-Item Measures of Doubly Concrete Constructs." *International Journal of Advertising* 28(4):607–21. doi:10.2501/S0265048709200783.
- Cho, Eunseong and Seonghoon Kim. 2015. "Cronbach's Coefficient Alpha: Well Known but Poorly Understood." *Organizational Research Methods* 18(2):207–30. doi:10.1177/1094428114555994.
- Cohn, Norman. 1993. *Europe's Inner Demons: The Demonization of Christians in Medieval Christendom*. London, UK: Random House.
- Comrey, Andrew L. and Howard B. Lee. 2013. *A First Course in Factor Analysis*. New York, NY: Psychology Press.
- Cronbach, Lee J. 1951. "Coefficient Alpha and the Internal Structure of Tests." *Psychometrika* 16(3):297–334. doi:10.1007/BF02310555.
- Dawes, John. 2008. "Do Data Characteristics Change According to the Number of Scale Points Used? An Experiment Using 5-point, 7-point and 10-point Scales." *International Journal of Market Research* 50(1):61–104. doi:10.1177/147078530805000106.
- Desmond, Scott A., Tom Clark, and Christopher D. Bader. 2023. "Sympathy for the Devil: Belief in Satan and Moral Beliefs." *Deviant Behavior* 44(5):752–67. doi:10.1080/01639625.2022.2092786.
- Douglas, Jack D. 2017. *Everyday Life: Reconstruction of Social Knowledge*. Oxford, UK: Routledge.
- Ellison, Christopher G., Benjamin Dowd-Arrow, Amy M. Burdette, Pablo E. Gonzalez, Margaret S. Kelley, and Paul Froese. 2021. "Peace through Superior Firepower: Belief in Supernatural Evil and Attitudes toward Gun Policy in the United States." *Social Science Research* 99:102595. doi:10.1016/j.ssresearch.2021.102595.
- Engelhard, Georg Jr and Jue Wang. 2021. *Rasch models for solving measurement problems: Invariant measurement in the social sciences*. London, UK: Sage.
- Evans, Gillian R. 1982. *Augustine on Evil*. Cambridge, UK: Cambridge University Press.
- Froese, Paul and Christopher D. Bader. 2010. *America's Four Gods: What We Say about God—and What That Says about Us*. Oxford, UK: Oxford University Press.
- Fuchs, Christoph and Adamantios Diamantopoulos. 2009. "Using Single-Item Measures for Construct Measurement in Management Research: Conceptual Issues and Application Guidelines." *Die Betriebswirtschaft* 69(2):195–210.
- Gardner, Donald G., L. L. Cummings, Randall B. Dunham, and Jon L. Pierce. 1998. "Single-Item Versus Multiple-Item Measurement Scales: An Empirical Comparison." *Educational and Psychological Measurement* 58(6):898–915. doi:10.1177/0013164498058006003.

- Gierl, Mark J., Jeffrey Bisanz, Gay L. Bisanz, Keith A. Boughton, and Shameem Nyla Khaliq. 2001. "Illustrating the Utility of Differential Bundle Functioning Analyses to Identify and Interpret Group Differences on Achievement Tests." *Educational Measurement: Issues and Practice* 20(2):26–36. doi:10.1111/j.1745-3992.2001.tb00060.x.
- Gorsuch, Richard L. and Sam G. McFarland. 1972. "Single vs. Multiple-Item Scales for Measuring Religious Values." *Journal for the Scientific Study of Religion* 11(1):53–64. doi:10.2307/1384298.
- Gorsuch, Richard L. and Susan E. McPherson. 1989. "Intrinsic/Extrinsic Measurement: I/E-Revised and Single-Item Scales." *Journal for the Scientific Study of Religion* 28(3):348–54. doi:10.2307/1386745.
- Grasmick, Harold G. and Anne L. McGill. 1994. "Religion, Attribution Style, and Punitiveness toward Juvenile Offenders." *Criminology* 32(1):23–46. doi:10.1111/j.1745-9125.1994.tb01145.x.
- Greiff, Samuel and Mark S. Allen. 2018. "EJPA Introduces Registered Reports as New Submission Format." *European Journal of Psychological Assessment* 34(4):217–19. doi:10.1027/1015-5759/a000492.
- Jung, Jong H. 2020. "Belief in Supernatural Evil and Mental Health: Do Secure Attachment to God and Gender Matter?" *Journal for the Scientific Study of Religion* 59(1):141–60. doi:10.1111/jssr.12645.
- Leiber, Michael J. and Anne C. Woodrick. 1997. "Religious Beliefs, Attributional Styles, and Adherence to Correctional Orientations." *Criminal Justice and Behavior* 24(4): 495–511. doi:10.1177/0093854897024004006.
- Lemert, Edwin M. 1997. *The Trouble with Evil: Social Control at the Edge of Morality*. New York, NY: State University of New York Press.
- Linacre, John Michael. 2002. "What Do Infit, Outfit, Mean-Square, and Standardization Mean?" *Archives of Rasch Measurement* 16:871–82.
- Loo, Robert. 2002. "A Caveat on Using Single-Item Versus Multiple-Item Scales." *Journal of Managerial Psychology* 17(1):68–75. doi:10.1108/02683940210415933.
- MacCallum, Robert C., Keith F. Widaman, Kristopher J. Preacher, and Sehee Hong. 2001. Sample Size in Factor Analysis: The Role of Model Error. *Multivariate Behavioral Research* 36(4):611–37. doi:10.1207/S15327906MBR3604\_06.
- Martinez, Brandon C. 2013. "Is Evil Good for Religion? The Link between Supernatural Evil and Religious Commitment." *Review of Religious Research* 55(2):319–38. doi:10.1007/s13644-012-0094-x.
- Martinez, Brandon C., Joshua C. Tom, and Joseph O. Baker. 2022. "Flowing across with Demonic Hate: Belief in Supernatural Evil and Support for Stricter Immigration Policy." *Journal for the Scientific Study of Religion* 61(2):408–32. doi:10.1111/jssr.12787.
- Martinez, Brandon C., Joshua C. Tom, Todd W. Ferguson, Brita Andercheck, and Samuel Stroope. 2018. "Parenting Practices and Attitudes and the Role of Belief in Supernatural Evil: Results from a National U.S. Survey." *Journal of Family Issues* 39(6):1616–38. doi:10.1177/0192513X17720757.
- Matthews, Russell A., Laura Pineault, and Yeong-Hyan Hong. 2022. "Normalizing the Use of Single-Item Measures: Validation of the Single-Item Compendium for Organizational Psychology." *Journal of Business and Psychology* 37(4):639–73. doi:10.1007/s10869-022-09813-3.
- Messadié, Gerald. 1996. *A History of the Devil*. New York, NY: Kodansha International.
- Pietryka, Matthew T., Randall C. MacIntosh. 2022. ANES Scales Often Do Not Measure What You Think They Measure. *The Journal of Politics* 84(2):1074:1090. <https://doi.org/10.1086/715251>.

- Pagels, Elaine. 1996. *The Origin of Satan*. New York, NY: Vintage.
- Ray, John J. 1974. "Are Trait Self-Ratings as Valid as Multi-Item Scales? A Study of Achievement Motivation." *Australian Psychologist* 9(1):44–49. doi:10.1080/00050067408256497.
- Robitzsch, Alexander, Thomas Kiefer, and Margaret Wu. 2024. *TAM: Test Analysis Modules*. R package version 4.2-21. Retrieved (<https://CRAN.R-project.org/package=TAM>).
- Rossiter, John R. 2002. "The C-OAR-SE Procedure for Scale Development in Marketing." *International Journal of Research in Marketing* 19(4):305–35. doi:10.1016/S0167-8116(02)00097-6.
- Russell, Jeffrey B. 1981. *Satan: The Early Christian Tradition*. New York, NY: Cornell University Press.
- Russell, Jeffrey B. 1992. *The Prince of Darkness: Radical Evil and the Power of Good in History*. New York, NY: Cornell University Press.
- Schweizer, Karl. 2020. "Method Effects in Psychological Assessment." *Psychological Test and Assessment Modeling* 62(3):337–43.
- Sigelman, Lee. 1977. "Multi-Nation Surveys of Religious Beliefs." *Journal for the Scientific Study of Religion* 16(3):289–94. doi:10.2307/1385699.
- Smith, Tom W. 1998. "A Review of Church Attendance Measures." *American Sociological Review* 63(1):131–36. doi:10.2307/2657485.
- Smith, Adam B., Robert Rush, Lesley J. Fallowfield, Galina Velikova, and Michael Sharpe. 2008. "Rasch Fit Statistics and Sample Size Considerations for Polytomous Data." *BMC Medical Research Methodology* 8:33. doi:10.1186/1471-2288-8-33.
- Wanous, John P. and Arnon E. Reichers. 1996. "Estimating the Reliability of a Single-Item Measure." *Psychological Reports* 78(2):631–34. doi:10.2466/pr0.1996.78.2.631.
- Wilcox, Clyde, Sharon Linzey, and Ted G. Jelen. 1991. "Reluctant Warriors: Premillennialism and Politics in the Moral Majority." *Journal for the Scientific Study of Religion* 30(3):245–58. doi:10.2307/1386971.
- Wilson, Kenneth M., Jennifer L. Accord, and Ronan S. Bernas. 2006. "Intolerance toward Others and Belief in an Active Satan." *Research in the Social Scientific Study of Religion* 17:39–48. doi:10.1163/9789047411413\_006.
- Wilson, Kenneth M. and Jennifer L. Huff. 2001. "Scaling Satan." *The Journal of Psychology* 135(3):292–300. doi: 10.1080/00223980109603699.
- Wolff, Kurt H. 1969. "For a Sociology of Evil." *Journal of Social Issues* 25(1):111–26. 10.1111/j.1540-4560.1969.tb 02581.x.

### Author biographies

**Tom Clark** is a Lecturer in Research Methods at the University of Sheffield. He is interested in all aspects of research method and methodology, particularly with respect to learning and teaching. His other interests have variously focused on the sociology of evil, student experiences of higher education, and soccer fandom.

**Scott A. Desmond** is an Assistant Professor of Sociology in the Division of Liberal Arts at Indiana University Purdue University-Columbus. He received his Ph.D. in Sociology from the University of Washington. Although his research focuses primarily on how religion is related to delinquency, substance use, and subjective well-being, he

also studies how neighborhood characteristics and low self-control contribute to crime, delinquency, and substance use.

**Christopher D. Bader** is a Professor of Sociology at Chapman University and Campbell Professor of Behavioral Sciences. He is affiliated with the Institute for Religion, Economics, and Society (IRES) and is co-director of the annual Chapman University Survey of American Fears. He is associate director of the Association of Religion Data Archives (ARDA), funded by the Templeton Foundation and Lilly Foundation. Christopher is the author of five books and over 50 articles and chapters in the fields of sociology, deviance, criminology, the sociology of religion, and education.