Assessing the Relationship Between Supernatural Punishment and Large-scale Cooperation Through Causal Modelling

MA thesis Olivier van Baars

24,458 words

Faculty of Theology & Religious studies

Primary supervisor: Gorazd Andrejč

Secondary assessor: Benjamin Purzycki

Table of Contents

1. Introduction

One of the great mysteries of human cultural history is the upscaling of small-scale societies to large-scale cooperative societies (Norenzayan et al., 2016; Powers et al., 2021; Henrich, 2016; Turchin et al., 2022). For much of human history, groups of humans used to live together in small hunter-gatherer societies not exceeding around 150 people (Henrich, 2016). At such scales, cooperation is stable. These small-scale hunter-gatherer societies are characterized by prosocial behaviors, such as food sharing. It is common that such cooperative behaviors are reciprocated in small groups consisting of kith and kin (Axelrod and Hamilton, 1981; Henrich, 2016). Furthermore, the costs of cheating and defecting are high, as in such small groups everyone knows each other personally or was a family member – being caught cheating would be detrimental to reputation.

Then, around 12,000 years ago, the advent of agricultural technology enabled an increased food production that was capable of sustaining larger groups and societies (Bocquet-Appel, 2011). This agricultural revolution caused a major demographic shift known as the Neolithic Demographic Transition (ibid.). Larger-scale societies and groups are characterized by sustained cooperation and trust in anonymous strangers. It is unclear whether cooperative behavior is reciprocated in such impersonal relationships. Cooperation at this scale enhances a phenomenon called free-riding, which means that one benefits from the cooperative behaviors of others without contributing or reciprocating. This is because the cost of cheating and defecting in small-scale societies is higher than in larger-scale societies. Consequently, large-scale societies would not be socially stable (Chudek and Henrich, 2011).

However, large-scale cooperation is extremely common in our species, and it has been the key factor in the global ecological domination of humans (Henrich, 2016). Therefore, this phenomenon requires an explanation: what is capable of generating and sustaining large-scale cooperation? Some have argued that current theories of kinship and reciprocal altruism are incapable of explaining large-scale cooperation (Norenzayan et al., 2016; Chudek and Henrich, 2011). In the last two decades, evolutionary scholars of religion have theorized how vigilant, powerful, and punitive deities may have contributed to stabilizing and stimulating large-scale cooperation (Johnson, 2015; Norenzayan et al., 2016). The foundational argument underlying these theories is that the looming threat of punishment of a supernatural deity raises the costs of free-riding in such a way that it stimulates cooperation over defection or cheating (Johnson and Krüger, 2004; Johnson, 2015). Could particular features of religion explain why large-scale societies are so socially stable and ecological successful? In order to answer this question, we first need to understand the causal structure underlying the relationship between supernatural punishment and cooperation. Hitherto, this step has not been undertaken and is the central goal of this MA thesis.

1.1. Problem statement and purpose of study

Recently, scholars have disagreed about the nature and assessment of the causal relationship between supernatural punishment and cooperation (Turchin et al., 2022; Watts et al., 2015). There are multiple causal pathways to both supernatural punishment and large-scale cooperation, which requires unfolding the complex causal web of variables involved (Norenzayan et al., 2016). Additionally scholars also use diverging operational constructs that refer to supernatural punishment as well as large-scale cooperation. Another area of contention revolves around methodology and data. Much research on the topic relies on old ethnographic records collected by missionaries and ethnographers over a century ago (Watts et al., 2022). These records have been shaped by certain biases, outdated theoretical assumptions, and missing data. For this reason, some scholars are skeptical of making strong inferences based on this data (ibid.). Disagreements about the nature of the causal relationships have led to lively debates about what the best way is to proceed in testing hypotheses that connect supernatural punishment to cooperation (Turchin et al., 2022; Slingerland et al., 2020; Watts et al., 2015; Purzycki et al., 2022).

To contribute to this discussion, this thesis identifies three different causal models regarding the nature of the relationship between supernatural punishment and cooperation. These are: 1) the supernatural punishment hypothesis-model (SPH), developed by Johnson and Krüger (2004); 2) the prosocial religions model developed by Norenzayan et al., (2016); 3) the socio-ecological model (e.g., Purzycki et al., 2022). The SPH model argues that supernatural punishment increases cooperation by virtue of a punitive deity increasing the costs of defecting and cheating. The prosocial religions model argues that cooperation and supernatural punishment co-evolve, meaning that large-scale cooperation contributes to the spread of punitive agents and that these agents in turn stimulate cooperation at increasing scales. The socio-ecological model considers supernatural punishment as an evolved response to local socio-ecological challenges, such as climate stability, resource scarcity, warfare as well as large-scale cooperation. Each of these models compounds a different causal structure.

Although such causal hypotheses revolving around the relationship between supernatural punishment and cooperation are common, research in the field has rarely effectively employed causal methodologies (Bulbulia et al., 2021; Watts et al., 2016; Purzycki et al., 2022). This is highly problematic, as drawing causal inferences without properly considering the causal structure of the research problem may lead to biased or wrong conclusions (McElreath, 2020).

There are many benefits to taking a causal approach. One of them is that it requires researchers to spell out their causal assumptions that derive from theory (McElreath, 2020). Another is that a causal approach visualizes the potential confounding relationships that need to be held constant during analysis. If confounders are not held constant, this may change the conclusions of the

study. As few studies have taken an explicit causal approach, this thesis employs causal analysis to assist in mitigating this issue. I use procedures devised in causal inference theory, such as simulating from directed acyclic graphs (e.g., McElreath, 2020), that assist in disentangling the causal constellation of variables involved.

In sum, a major limitation in this debate is that an explicit causal approach to the relationship between supernatural punishment and cooperation is lacking. Most studies in this area are correlational or predictive, while few successfully apply causal inference (Watts et al., 2015; Watts et al., 2016). The foremost aim of this thesis is to contribute to this scientific debate by carefully deconstructing the diverging causal claims, and in so doing, I suggest how we may go about in testing different causal hypotheses regarding the relationship between supernatural punishment and cooperation. Additionally, I critically examine various issues across all levels of scientific inquiry, including data, methodologies, concepts and theory. To address these issues, I suggest theoretical, conceptual, and causal workflows based on the philosophy of psychology and causal inference theory.

1.2. Structure of thesis

This thesis consists of five chapters, the first being the introduction.

The second chapter reviews the theoretical background of the debate. I outline the different claims that have been made about the causal nature of the relationship between supernatural punishment and cooperation. The postulated causal relationships in the literature will form the backbone of my causal analysis. Then, I discuss issues related to concepts and their operationalizations. Here, I argue that careful consideration of the lucidity and transparency of the key concepts and operationalizations is pivotal for theoretical progress.

The third chapter contains an introduction to causal inference theory before proceeding to the causal analysis of the supernatural punishment-cooperation relationship. The purpose here is to substantiate causal structures, which show what variables are required in the analysis to make valid causal inferences. The secondary purpose is to show what will happen to the result if confounding variables are improperly considered in the statistical analysis. In doing so, I show how and why a strong causal approach is necessary to move the field forward.

In the fourth chapter, I consider how methodological challenges related to cross-cultural database design, Galton's problem, ethnographic data quality, and causal inference have shaped the debate. I provide suggestions on how to address these challenges.

The final chapter contains a discussion on how the field may proceed from here. I integrate recent literature on the philosophy of psychology which may be applied to future evolutionary

research on religion. Based on my inquiry into the issues revolving around data, methodologies, concepts and theories, I provide general recommendations for the field. Additionally, I argue that adopting a wider range of economic games in ethnographic studies may be conducive to deepening our understanding of the relationship between supernatural punishment and cooperation. Finally, I extend these recommendations out to the other social sciences.

1.3. Research Questions

RQ: What are the causal structures of the models explaining the link between supernatural punishment and cooperation?

SQ1: What is the causal structure of the supernatural punishment hypothesis?

SQ2: What is the causal structure of the prosocial religions model?

SQ3: What is the causal structure of the socio-ecological model?

SQ4: What do these causal models imply for future empirical research in the evolutionary science of religion?

2. Background

This chapter reviews the main perspectives and challenges that arise when investigating the relationship between supernatural punishment on large-scale cooperation. I first sketch the theoretical landscape before moving on to conceptual and operational challenges.

2.1. Theoretical background

If we wish to understand the causal relationship between supernatural punishment and large-scale cooperation, we need to untangle the complex causal constellation of variables involved. To do so, this section is devoted to dissecting the different theoretical claims that have been made regarding the nature of supernatural punishment and large-scale cooperation. I discuss perspectives from cultural evolutionary theory, human behavioral ecology as well as claims relating to the existential security hypothesis and life-history theory.

Of course, both supernatural punishment and cooperation are complex, multifaceted concepts, and a full multi-level account of their causal structure is beyond the scope of this thesis. According to Tinbergen's classical framework for explaining behavior, there are at least four perspectives for explaining behavior (1968). These concern: development, mechanism, evolution, and function. In this thesis, I am mainly concerned with the functional aspects of supernatural punishment

and cooperation — that is, the potential adaptive consequences and persistence of the trait. However, to understand these consequences, it is also important to understand *how* and *why* (mechanism) supernatural punishment potentially elicits prosocial tendencies. Therefore, I mainly confine myself to these two aspects.

2.1.1. Supernatural punishment and cooperation

Supernatural punishment generally assumes a central position in contemporary evolutionary theories of religion. For example, the work of Johnson (2005, 2015) argues that supernatural punishment plays a key role in explaining cooperation at both small and large scales, given that 'religious traditions, taboos and mythology provide […] the norms of conduct'; Johnson and Krüger, 2004, 163). The work of Norenzayan (2013) and associates (2016) also reserves a central position for supernatural punishment, although they connect it to other aspects of the religious 'package' (such as ritual, costly, signals, and credibility enhancing displays). Additionally, their emphasis is more on specific kinds of supernatural agents (called Big Gods, that are powerful and omniscient) who perform the punishment.

To pinpoint the key differences between the two frameworks, Schloss and Murray (2011) argue that the Supernatural Punishment Hypothesis (hence: SPH) can be distinguished on the basis of the mechanism by which cooperation is stimulated. One is referred to as 'Punishment Avoidance' and the other as 'Cooperation Enhancement'. The former postulates that supernatural punishment leads to individual-level adaptive advantage due to individuals not having their cheating punished. This version of the SPH is primarily expounded in the work of Johnson (2005, 2015), Johnson and Bering (2006,) and Johnson and Krüger (2004). Supernatural punishment increase cooperation due to the potential punishment incentivizing adherence to prosocial norms.

As for the Cooperation Enhancement account, primarily advocated for by Norenzayan and others (2016), supernatural punishment enhances cooperation and deters defection in situations where most other mechanisms that stabilize large-scale prosociality unwind. Large-scale cooperation can be achieved if the costs of potentially being punished outweigh the costs of cooperating. Other punishment systems (such as the police) are themselves costly, requiring people to contribute valuable time and resources to them as well as risk retribution from those punished (Johnson, 2015). This entails that the development of such punishment systems are themselves subject to free riding, called second-order free riding (Ozono et al., 2016). Finally, such punishment systems are only as effective as their way of detecting cheating. If cheating cannot be detected in some cases, this will be exploited by cheaters when they realize that is the case. Therefore, supernatural punishment appears to be the silver bullet that overcomes all of these problems. Watchful and vigilant gods are generally always present and monitoring, which makes for an excellent cheating detection mechanism (Johnson, 2005). Furthermore, it circumvents the issue of second-order free riding completely, as

belief in supernatural punishment is costless (Johnson and Krüger, 2004).

As Johnson (2015) puts it, these two versions of the SPH are not mutually exclusive. They merely place different emphases. Johnson (ibid.) argues that his Punishment Avoidance account is about explaining the origins of religion in the Pleistocene (c. 2.5 million years ago to c. 12,000 years ago). It is primarily conceived in terms of individual-level and genetic selection. In turn, the Cooperation Enhancement account focusses more on the emergence of large-scale societies in the Holocene (c. 12,000 years ago to the present day) and is primarily conceived of in group-level and cultural selection terms.

Some are critical that supernatural punishment functions as a silver bullet capable of explaining cooperation. For example, Murray (2009) notes that whenever an individual cheats without receiving any supernatural punishment, they will realize that such punishment may not even exist, leading to that system breaking down. Although that might be true for some individuals, as long as most people in the group continue to belief in supernatural retribution, cooperation will remain stable at larger scales – no system is watertight and every kind of social organization will feature cheats (Johnson, 2015). What matters for evolution is that belief in supernatural punishment raises the potential for survival during times of competition. This is where contextual factors come into play. For example, Schloss and Murray (2011) argue that the supernatural punishment mechanism may be stabilized by other features of religious systems. Such features may involve karmic beliefs or a deferral of punishment to the afterlife (Johnson, 2015). Therefore, supernatural punishment can be fruitfully posited as one potential mechanism capable of sustaining and enabling large-scale cooperation.

Johnson (2015) and Norenzayan et al.'s (2016) frameworks also differ in terms of causal structure. Johnson's framework is arguably more straightforward, as he simply argues that supernatural punishment can increase (large-scale) cooperation. Additionally, Johnson and Bering (2006) reserve also a role for supernatural reward, although they emphasize that punishment is a more effective mechanism for engendering cooperation than reward (223, their italics):

> The effects of carrots [reward] and sticks [punishment] on the level of cooperation are not symmetrical, even when of equivalent magnitude: punishment is inherently *more* effective at promoting cooperation than rewards. Carrots are not enough because, although they may encourage *some* people to cooperate, they do not prevent *all of them* from cheating. Even if the rewards of cooperation are large and obvious to everyone involved, they provide no credible deterrent against defectors.

Beside supernatural reward, Johnson (2005) argues that punishment by other humans are also capable of stimulating cooperation. Additionally, a potential lack of a human punishment system (e.g., police force) creates a selection pressure for the development of supernatural punishment. Although

Johnson does not include variables on human punishment or supernatural reward, he suggests that future empirical work ought to do so.

Norenzayan et al.'s (2016) prosocial religions framework is a lot more complex, as it seeks to both explain the rise of prosocial religions (and not just supernatural punishment) as well as the rise of large-scale societies (and not just cooperation). They argue that increased food production following the agricultural revolution increased group sizes, leading to higher selection pressures for mechanisms capable of stabilizing large-scale cooperation. Additionally, their framework relies on the existential security hypothesis, which argues that beliefs in supernatural punishment may be stimulated through resource scarcity, and intergroup-competition (such as warfare). Finally, their framework emphasizes that prosocial religions and large-scale societies co-evolve in a feedback loop, such that prosocial religions contribute to large-scale cooperation and that large-scale cooperation in turn increases the spread and development of those religions. However, they note that their model is not a complete account of the complex relationship between religion and society (3):

> There are several other cultural evolutionary paths to large-scale cooperation, including institutions, norms, and practices unrelated to prosocial religions. These include political decision making (e.g., inherited leadership positions), social organization (e.g., segmentary lineage systems), property rights, division of labor (e.g., castes), and exchange and markets. The causal effects of religious elements can interact with all of these domains and institutions, and this causality can run in both directions, in a feedback loop between prosocial religions and an expanded cooperative sphere

I now proceed to discuss evidence addressing the supernatural punishment and cooperation relationship in laboratory experiments, ethnographic fieldwork, and cross-cultural database studies.

Studying supernatural punishment in experimental and laboratory settings often involves priming religious concepts to participants. Priming religious concepts, such as gods, generally reduces cheating and selfish behavior, while promoting cooperation and fairness towards anonymous strangers (Norenzayan, 2013). For example, Shariff and Norenzayan (2007) demonstrate that individuals in anonymous economic games were more likely to donate money to strangers when god concepts were primed. Moreover, there appears to be a distinction between whether a supernatural agent is punitive or benevolent, as predicted by Johnson and Bering's framework (2006). Shariff and Norenzayan (2011) find that belief in punitive gods strongly reduces moral transgressions, such as cheating. However, if participants believed that the god was benevolent, this increased cheating behavior. Another priming study finds that supernatural beliefs increase feelings of being socially monitored (Gervais & Norenzayan, 2012). Therefore, if people believe that they are being constantly

monitored and policed, this will reduce norm transgressions and cheating. Furthermore, reaction-time analyses conducted by Purzycki et al. (2012) found that people respond faster to questions about punitive gods' knowledge about norm-transgressing behavior than about other kinds of behavior. In addition, Laurin et al. (2012) conducted five studies using a mix of economic games and surveys to find that belief in a punitive god leads to less punishing behavior towards free-riders, adding support to the idea that the costs of punishment are indeed offloaded to a supernatural agent. In summary, these experimental studies provide broad support for supernatural punishment influencing prosocial and cooperative behavior, although this branch of research is limited by a lack of cross-cultural comparison.

Aside from social psychological studies in experimental settings, much research on supernatural beliefs are conducted in anthropological field settings. Ethnographic research on supernatural punishment typically involves hybrid methodologies that employ both traditional anthropological research methods such as interviewing and participant observation, as well as more experimental setups involving economic games. The Evolution of Religion and Morality project is a good example of research employing such methods (Purzycki et al., 2022; Purzycki et al., 2018; Lang et al., 2019). A major advantage of using economic games is that they have been validated and reflect real-world behavior and have a reputable, widely-used track record in behavioral economics and cross-cultural research (Pisor et al., 2020).

To examine whether supernatural punishment extends cooperation to geographically distant co-religionists, Lang et al. (2019) collected data among 15 diverse societies exercising different modes of subsistence. One of these games was a modified Random Allocation Game (RAG). Here, participants are to distribute 30 coins over two cups. In the first version of the game, one cup was for the participant, and the other cup for a distant co-religionist. In the second version of the game, one cup was for a local co-religionist and the other for a distant co-religionist. The participants play in private and use a die with three sides of one color and three sides of another color. Before each dice roll, the participants pick a cup (mentally). Subsequently, participants are instructed to put the coin into the cup respective to the color that comes up on the die. The crux of the matter is that participants can of course choose whatever cup they like; there only is mental pressure to pick the corresponding cup.

The other economic game was the Dictator Game (DG). Here, participants distribute 10 coins over two cups (in a similar setup as the RAG) in whatever way they like. The researchers find that higher rates of beliefs of gods as monitoring or punitive is related to higher resource sharing with distant co-religionists. Resource sharing with anonymous strangers is considered an important aspect of large-scale cooperation, as it is a costly interaction with a non-kin stranger for which the potential benefit is opaque. However, the researchers also found that supernatural agents in small-scale

societies are often restricted to local ecologies, making it hard to establish a geographically-overarching identity between local and distant co-religionists.

However, a recent study examining free-listed data on whether gods' explicitly formulated moral concerns predicts cooperative play found no clear association between such concerns and cooperation in the same 15 societies, although the scholars did find an association between punishment and cooperation (Bendixen et al., 2023). In another recent study combining both online and on-site methods, Pasek et al. (2023) find that when people are primed to think about their god, they increase the amount of money they share with people from religious outgroups, contra to Norenzayan et al.'s (2016) parochial prosociality hypothesis. However, their sample was confined mostly to Abrahamic religions. Henrich et al. (2010) used a similar approach by using a range of behavioral economic games on 15 different societies and found that community size is associated with (non-supernatural) punishment, such that larger communities have a higher presence of punishment. On the other hand, Ge et al. (2019) conducted donation games and dice allocation games in 17 communities varying in scale in China, but they did not find support for the idea that supernatural punishment nor community size predicted cheating in either economic game. Instead, they argue, reputational concerns are the most important motivator in these games. Overall, there is mild support for the influence of supernatural punishment on cooperative behavior. Ethnographic research employing behavioral economic experiments can provide fine-grained and theoretically validated methods to test the supernatural punishment hypothesis, but are limited in terms of large cross-cultural comparisons due to the resource intensity of gathering such data.

The final branch of literature employs large-cross-cultural databases to probe the relationship between religion and social life. The primary advantage of such databases is that they usually contain quantitative variables about religion, mode of subsistence, gender, and religion on hundreds of cultures, which enables world-wide systematic cross-cultural comparison. What discerns database studies from other ethnographic comparative research is that database studies are at the group level, while the ethnographic studies discussed in the previous section are all individual-level. This entails the potentially problematic assumption that cultural groups are considered homogenous in inferences. As such, all individual-level cultural variation is lost, which is problematic when there is a nontrivial proportion of atheism or extensive (religious) diversity present within a society (Ember and Ember, 2009).

The first large cross-cultural database study that investigated the link between religion and cooperation was spearheaded by Swanson (1960), who used a sample of 50 societies from the Ethnographic Atlas (EA). Based on Durkheimian sociological theories of religion, Swanson posited that supernatural beliefs developed as responses to the social structure of sovereign groups. The more

socially complex a society, the higher the presence of supernatural monitoring. His research was unsuccessfully replicated by Underhill (1975) and Davis (1971), who found that besides political complexity, economic complexity is a stronger predictor of monotheism. Swanson's findings were only partially replicated by Peregrine (1996), who investigated 72 native American societies. He found Swanson's cross-cultural methodology lacking, especially its unclear coding guidelines, yet did find some support for his original thesis. Roes and Raymond (2003), using the Standard Cross-Cultural Sample (SCCS), offer a similar line of reasoning to Underhill (1971), although based on the theoretical framework of moral systems biology (Alexander, 1987). This framework states that the pressure for social disjuncture increases as groups become larger due to inter-group competition over habitat and food supply. Belief in moralizing gods then serves as a mechanism to bind the group together by imposing a set of moral rules and norms. Similarly, Johnson (2005) also used the SCCS as a test of the SPH. He found largely similar results to Roes and Raymond, but the SPH puts forward the reverse causal association, namely that supernatural punishment is not a response to large-scale cooperation, but rather a cause of it.

However, the statistical analyses conducted by Roes and Raymond and Johnson are merely correlational and are therefore incapable of teasing apart the causal direction (Watts et al., 2015). This only highlights the demand for a strong causal approach to this debate. This is what Watts and associates (2015) set out to do by conducting a Bayesian phylogenetic analysis of Polynesian cultures. This method is able to tease apart the causal and temporal ordering of how certain cultural traits evolved. They argue that supernatural punishment by a broad range of moralizing agents drove political complexity.

More recently, Turchin et al. (2022) and Whitehouse et al. (2022) cast doubt on these claims, arguing that supernatural punishment has no causal influence on socio-political complexity whatsoever. To test this, they constructed a purpose-built large-scale cross-cultural database called Seshat containing diachronic data on cultures, designed to directly reflect the potential cultural evolution of societies. Their analyses suggest that moralizing gods and punishment began to appear only after societies had already upscaled. Additionally, based on existential security hypothesis, they argue that both social complexity and supernatural punishment evolved as a response to warfare and agricultural productivity, creating a spurious association between supernatural punishment and social complexity. However, their approach has been criticized both from historical perspectives (Slingerland et al., 2020; although see Larson et al., 2022) as well as methodological and causal perspectives (Beheim et al, 2021; Purzycki et al., 2022). This led to a retraction due to inappropriate handling of missing data (Whitehouse et al., 2019; Whitehouse et al., 2021; Beheim et al., 2021).

In sum, there appears to be broad support for the supernatural punishment hypothesis,

ranging from laboratory experiments to ethnographic field work to large-scale database analyses. That said, Norenzayan et al. (2016) note that there are multiple pathways that lead to large-scale cooperation. In addition, alternative theoretical frameworks suggest different causal structures concerning the relationship between supernatural punishment and large-scale cooperation. I now turn to one of these: the socio-ecological model.

2.1.2. Supernatural Punishment as an evolved response

A broad evolutionary literature has emerged in the last decade that argues that religious beliefs can be considered evolved responses to socio-ecological circumstances and slow life history strategies (Baumard and Boyer, 2013; Baumard et al., 2015; Purzycki et al., 2022). In addition, the existential security hypothesis posits that the perceived vulnerability to personal and social threats positively correlates to belief in moralizing gods in the absence of secular institutions (Baimel et al., 2021). As such, important factors that influence religious beliefs are resource abundance, war, and ecological circumstances such as prevalence of drought, rainfall, and natural disasters (Botero et al., 2014; Turchin et al., 2022). This section spells out the proposed causal mechanisms of how such existential threats and local socio-ecological problems potentially influence the development of supernatural punishment beliefs.

Human behavioral ecology accounts for religious variation by arguing that such variation is a direct reflection of local socio-ecological challenges (Purzycki et al., 2022) Furthermore, religious systems appear to directly address and adapt to these challenges (Purzycki and Sosis, 2022). For example, a society constantly plagued by droughts is more likely to have a god addressing these droughts. Therefore, in our case, religious systems may offer solutions to the social challenge of large-scale cooperation. Subsequently, religious systems that directly address the challenge of large-scale cooperation yield adaptive advantages for the people that adhere to them (Purzycki et al., 2022).

One particular perspective from life history theory offers that certain features of large-scale societies, such as affluence and material security, contributed to the development of moralistic religions (Baumard et al., 2015). This is due to increases in affluence affecting human motivation and reward systems, causing humans to move from ‘fast’ to ‘slow’ life history strategies. Fast strategies are characterized by short-term investment of resources, which includes early reproduction and coercive interactions, while slow strategies are characterized by less materialistic goals and cooperative interactions (ibid.). Often, changes in the environment, such as increased droughts or natural disasters, engender a shift in life history strategy. The claim is that moralistic religions featuring supernatural punishing agents that emphasize these slow life history strategies are thus reflections of increased affluence and wealth (ibid.). This life history model therefore predicts that more materially

secure and affluent societies and cultural groups will attribute greater moral concern to supernatural agents. However, Purzycki et al. (2018) found no support for such a relationship in their cross-cultural study. Their individual-level study finds no relationship between number of children, material security, or formal education (which are indicators of a slow life-history strategy) and religious beliefs and behaviors. As such, the life-history model is contended.

The existential security hypothesis also makes predictions about the influence of ecological circumstances, such as the prevalence of natural disasters, drought, and resource abundance on both supernatural punishment beliefs and large-scale cooperation. Therefore, these variables are potential confounders in determining the relationship between supernatural punishment and cooperation. There is widespread support in the literature that such environmental cues increase religious beliefs. For example, Sibley and Bulbulia (2012) found increased commitment to moralistic religions after an earthquake in New Zealand.

However, there is mixed evidence on whether resource scarcity is positively or negatively associated with beliefs in moralistic gods. For example, Skoggard et al. (2020) find that resource stress predicts the involvement of moralizing gods with the weather and food supply, yet they find no support that such gods have significant effects on resource sharing. Using water scarcity as a proxy for resource stress, Snarey (1996) found that in water-abundant societies, gods were less concerned with human cooperation than in water-scarce societies. However, Roes and Raymond (2003) find the same relationship as Snarey, but they find no association between drought and belief in moralizing gods. To attest this, Botero et al. (2014) conducted a large cross-cultural study using the SCCS. They developed two composite variables: one indicating resource abundance (composed of rainfall abundance, primary production, and greater biodiversity) and the other climate stability (composed of exposure to predictable annual cycles of precipitation and temperatures). Their results indicate that beliefs in moralizing gods increases in times of ecological hardship, such as during droughts. Such beliefs can in turn improve a group's ability to cope with such circumstances by promoting cooperation and reducing existential anxiety. A similar study conducted by Spicer et al. (2022) used the Database for Religious History found no reliable relationship between those ecological variables and religious beliefs in supernatural punishment. However, the researchers argue that this may have been due to their study relying on inaccurate climate reconstructions. In sum, commitment to moralistic traditions could be considered an ecological adaptation, as groups featuring such traditions have a higher chance of surviving.

Another way to investigate the relationship between material security and religious beliefs is to consider a group's mode of subsistence. Usually, these modes are measured on a scale of agricultural productivity, which is typically divided into the categories of hunter-gathers,

horticulturalists, pastoralists, and agriculturalists. Increased agricultural productivity can facilitate population growth, increasing the demand of large-scale cooperation (Norenzayan et al., 2016). At the same time, stable food production predicts less belief in moralizing gods due to the increased material and existential security. However, large-scale food production may be susceptible to ecological threats, which would increase beliefs and commitments to moralizing gods. Peoples and Marlowe (2012) find that societies featuring pastoralism or agriculturalism have the highest prevalence of moralizing high gods, a finding corroborated by Turchin et al. (2022) and Underhill (1975).

The final prediction of the existential security hypothesis I discuss here is that warfare drives both religious beliefs and cooperation. Wars are situations that feature visceral instability in terms of personal and social vulnerability. Therefore, we would expect that warfare and other forms of intergroup competition increase religious commitments and beliefs. Henrich et al. (2019) find support for the idea that individuals seek out more religious commitments during war times. Furthermore, they find that exposure to war increases participation in religious organizations years after war. This supports the cultural evolutionary prediction that external threats – in this case, intergroup competition – increase adherence to social norms. Resultingly, war promotes cooperation directly as well as indirectly, mediated through increased belief and commitment to moralizing gods. Likewise, Turchin et al. (2022) also attribute a strong role of war in promoting both beliefs in moralizing supernatural punishment and socio-political complexity. However, their model posits no causal link between moralizing supernatural punishment and socio-political complexity, in contrast to the model of Henrich and others (2019).

Summarizing, I have identified three models of the causal relationship between supernatural punishment and cooperation. The first is Johnson and Krüger (2004), Johnson and Bering (2006) and Johnson's (2005, 2015) supernatural punishment hypothesis, which states that supernatural punishment increases cooperation on account of Punishment Avoidance. It is considered a genetic adaptation and focuses on individual-level selection in the Pleistocene. Norenzayan et al.'s (2016) framework is causally complex and also argues that supernatural punishment increases cooperation, but at the scale of larger societies in the Holocene at the level of the cultural group. The final model, the socio-ecological model, combines elements from life history theory, the existential security hypothesis, and human behavioral ecology. It postulates that supernatural punishment evolved as a response to socio-ecological challenges, including resource scarcity, intergroup competition, (large-scale) cooperation, and ecological stress.

To proceed, I discuss what conceptual and operational challenges have to be addressed to properly empirically engage with the predictions following these models.

2.2. Conceptual and Operational Issues

One persistent problem in the debate about the relationship between social complexity and moralizing religion revolves around operationalizing the key concepts. In other words, what do 'social complexity' and 'moralizing religion' actually refer to in an empirical setting? In attempting to assess the relationship between social complexity and moralizing religion, scholars have used many different operationalizations of them. This begs the question whether studies supposedly measuring the same concept are actually doing so. The phenomena 'religion', 'morality', and 'social complexity', are so sufficiently vague and ambiguous that they allow many distinct operationalizations (McKay and Whitehouse, 2015). As a result, it is hard to directly compare the results of the different studies conducted in this area. For example, Turchin et al. (2022) operationalize moralizing supernatural punishment in a different manner as opposed to e.g., Watts (2015). Yet, this fact is never addressed, even though they these studies attempt to contribute to the same scientific conversation. This is crucial, as one may reach radically different conclusions: for example, Brown (1991) argued that morals are universal, while Prinz (2007) claimed the opposite, which can be attributed due to them using different operationalizations of morality. This section develops the argument that these issues may be addressed by staying close to theory and explicitly addressing differences in operationalization, which may aid in increasing conceptual clarification (Bringmann et al., 2022). I first discuss issues pertaining to religion and morality, before moving on to challenges involving social complexity and large-scale cooperation.

2.2.1. Supernatural punishment and morality

Studies involving supernatural punishment involve a range of concepts relating to morality, cooperation and moralizing gods. Borrowing game theory methods from behavioral economics has been a popular choice to measure morality and cooperation concepts (e.g., Henrich et al., 2019; Lang et al., 2019). However, supernatural punishment does not have a generally accepted measure, which has caused problems in the past (e.g., by using the moralistic 'high' god construct), and may cause problems in the future (e.g., by lack of conceptual and operational clarification).

2.2.1.1. Morality, cooperation, and games

Anthropology has had troubles with developing a systematic concept of morality (Edel, 1962; Laidlaw, 2002). However, this has started to change recently in virtue of interdisciplinary collaborations with evolutionary biologists and behavioral economics (Curry et al., 2019; Pisor et al., 2020). Still, McKay and Whitehouse (2015) argue that studies that investigate the relationship between religion and morality suffer a multitude of problems. For example, the concepts religion and morality are often not properly broken down into theoretically grounded units. In the same vein, it is important that 'morality' should not be understood as in Western philosophy, but rather in terms of group-specific

norms that regulate social conduct (Lang et al., 2019). Such definitions that align closely to cooperation appear to be common in cultural evolutionary studies of religion (Bendixen et al., 2021).

This is the premise of the morality-as-cooperation account, developed by Curry (2016) and empirically assessed by Curry et al. (2019). It is rooted in game theory and argues that moral systems are bio-cultural solutions to the challenges posed by cooperation in day-to-day social life. Curry et al. review 7 types of cooperation and argue that each type explains a corresponding kind of morality, displayed in a table below (2019, 48).

Types of cooperation	**Types of corresponding morality**
Resource allocation to kin (kin altruism; Hamilton, 1963).	Family values
Coordination to mutual advantage (Lewis, 1969)	Group loyalty
Social exchange (Trivers, 1971)	Reciprocity
Conflict resolution through contests featuring displays of hawkish traits (Maynard Smith and Price, 1973)	Bravery
Conflict resolution through contests featuring displays of dovish traits (ibid.)	Respect
Division (Skyrms, 1996)	Fairness
Possession (Gintis, 2007)	Property rights

If supernatural punishment is about regulating social conduct and morality, we can hypothesize that gods will care about cooperative games. Resource allocation to kin revolves around caring for offspring and helping family members. As Curry (2016) notes, humans have invented various cultural institutions that extend the reach of kin altruism, such as naming conventions and incest-prevention rules. Religion may also be such an institution that prescribes those rules. Whenever such cooperative games form a natural selection pressure, we may expect that the local god punishes such behaviors.

Coordination to mutual advantage occurs when cooperation leads to more benefits than working individually. Curry (ibid.) suggests that humans are particularly good at this, importantly due to our theory of mind system, which allows us to infer the mental states of others. Furthermore, coalition forming as to compete with rival coalitions is also favored. The morality-as-cooperation account predicts that participation in collaborative endeavors, ingroup-favoritism and the adoption of local conventions will be regarded as morally good. Therefore, we might expect that the local gods

will punish behaviors not conducive to these aspects.

Social exchange refers to cooperative situations where mutualism is unclear. That is , benefits may not be clearly reciprocal and such cooperative games are subject to free riding. Typically in such games, non-cooperation is the only viable strategy. Such situations are also commonly labelled as prisoner's dilemma or public goods games. However, if such cooperative games occur repeatedly, the game turns into an 'assurance game', which will not lead to non-cooperation (Curry, 2016). The supernatural punishment hypothesis is usually centered around large-scale cooperative games, whereby such assurance is opaque, which would theoretically entail that non-cooperation would be the only viable strategy. This is the most researched kind of cooperative game in the supernatural punishment literature, and the general prediction is that the belief in gods that punish cheats in such games increase sharing behavior (e.g., Lang et al. 2019).

Conflicts present the opposing parties an opportunity to cooperate, as they may be able to compete in less mutually destructive ways. One way how cooperation may occur is for one party to flaunt their combative abilities and strength (called 'hawkish' displays; these displays are also commonly known as costly displays), scaring the other party to e.g., retreat or leave the resource the parties were competing over. Both groups gain from this, as they save on the potential costs of conflict. Conversely, submissive strategies (called 'dovish displays) may also be an attractive choice in already stable social groups in which the hierarchy of power is already present by virtue of e.g., reputation. Curry (2016) notes that this tendency for the strong to help the weak is cross-culturally widespread. Therefore, both hawkish and dovish traits will generally be morally favored, and resultingly, we can expect that gods will favor or punish such displays in various situations.

Division is concerned with how resources (such as from hunting or from borders between territories) should be divided, called 'bargaining problems' in game theory. Empirically, equal sharing is commonly cross-culturally attested in economic games (Henrich et al., 2005). Therefore, the model predicts that conflict over resources may be solved through division, explained by the moral virtue of fairness. As such, we can also expect that gods will punish non-fair behaviors.

The final cooperative situation in Curry's framework is that of possession. It revolves around how conflicts over resources can be resolved through recognition of prior ownership. This usually entails respecting the private property of others (and not stealing), be it resources or territory. Thus, we can expect that gods will punish stealing behaviors.

In the supernatural punishment literature, cooperation is usually construed in a fairly narrow manner, mainly focusing on issues revolving around social exchange. In other words, what we know about the relationship between religion and cooperation is confined to particular instances of cooperation (i.e., resource allocations). Research into the other kinds of cooperation as presented in

Curry et al.'s typology is less common. This begs the question if supernatural punishment only bears on cooperation in the sense of social exchange, or whether it extends to other possible forms of cooperation. Therefore, if we wish to understand the full scope of how supernatural punishment relates to cooperation, we will need to consider other cooperative games as well.

2.2.1.2. Moralistic (High) Gods and supernatural punishment

Early studies testing the supernatural punishment hypothesis primarily used the 'Moralizing High Gods' construct (e.g., Johnson, 2005; Roes and Raymond, 2003). The use of this construct could be considered an artefact of the availability of data and cross-cultural databases to test this idea (Lightner et al., 2023). The 'high god' variable in the SCCS is defined as 'a spiritual being who is believed to have created all reality and/or to be its ultimate governor, even though his/her sole act was to create other spirits who, in turn, created or control the natural world' (Murdock and White, 2006). This definition does not concern supernatural punishment directly – it only (presumably) implies it (Johnson, 2005). Furthermore, the SPH is agnostic to whether a supernatural agent is the creator of all reality because such characteristics are irrelevant to the punishment mechanism. The ubiquity of Moralizing High Gods in the literature thus reflects a practical choice rather than a theoretically motivated operationalization.

For this reason, evolutionary studies in the last decade have moved away from this construct by developing purpose-built databases and datasets (Watts et al., 2015; Lang et al., 2019; Turchin et al., 2022). Still, popular evolutionary frameworks of religion (e.g., Norenzayan et al., 2016) are theoretically informed by SCCS studies. Therefore, even though some frameworks, such as the prosocial religions account, have moved away from the moralistic high god-concept, the empirical work on which the framework is built still relies on that problematic 'high god' variable. Although recent studies by e.g., Watts and others (2015) are a welcome improvement, they still operationalize supernatural punishment in slightly different manners. They all intend to contribute to the same debate, yet there is little explicit dialogue on how these distinct operationalizations of supernatural punishment (and cooperation for that matter) have been constructed and how that may have affected the study.

Indeed, recent work in philosophy of science has argued that conceptual and operational clarification of that sort should be central to social scientific research (Bringmann et al., 2022). For example, Purzycki et al. (2016), Purzycki et al. (2018), and Lang et al. (2019) are concerned with 'moralizing gods', defined as 'belief in punitive and monitoring gods who care about norm transgressions and the Random Allocation Game (RAG) and Dictator Game (DG) allocations' (Lang et al., 2019, 4). Watts et al. (2015, 2) develop a somewhat similar operationalization, which they call

'Broad Supernatural Punishment', for which to be counted as present, 'there must be the concept of a supernatural agent or process that reliably monitors and punishes selfish actions, and this concept must (i) be widely advocated within the community, (ii) involve punishment of a broad range of selfish behaviours and (iii) apply to a wide range of community members'. On other hand, Turchin et al. (2022) develop an integrative measure of 7 variables related to supernatural punishment, which they call 'Moralizing Supernatural Punishment'. It includes whether supernatural punishment is predictable, targets individuals vs. the whole group, targets rulers, is subscribed to by the elites vs. the commoners, whether it occurs in the afterlife or in this life, and more. While this is more fine-grained, the authors do not theoretically motivate the deconstruction of the supernatural punishment concept. That aside, these studies all measure slightly different things, yet it is never addressed that this may influence the results. It is therefore essential that researchers (1) clearly theoretically motivate their operational construct, (2) address how their construct differs from those used in other studies and why, and (3) address how that construct could potentially influence the result of the study compared to other constructs (Bringmann et al., 2022).

2.2.2. Social complexity and large-scale cooperation

To estimate the influence of supernatural punishment on large-scale cooperation, scholars have used numerous operationalizations and adjacent proxy conceptualizations of cooperation, primarily social complexity. I will show why social complexity is a problematic concept. Furthermore, I provide a conceptual outline of cooperation and its relation to – often used interchangeably – the concepts of prosociality and altruism.

2.2.2.1. Measuring social complexity

While the MHG construct has received critical treatment in recent years (Purzycki and McKay, 2023; Lightner et al., 2023; Purzycki et al., 2016), social complexity has largely flown under the radar. This is striking, as they share a common intellectual history.

Using the EA, Swanson (1960) found that there is high positive relationship between high gods and 'political complexity', operationalized as the numbers of jurisdictional hierarchies beyond the local community of a society. The codes are: (1) no levels (no political authority beyond community); (2) one level (e.g., petty chiefdoms); (3) two levels (e.g., larger chiefdoms); (4) three levels (e.g., states); (5) four levels (e.g., large states). It should be noted that the exact coding criteria are unclear; it is not explicated what exactly constitutes a 'level of jurisdictional hierarchy', nor is it clear what distinguishes a chiefdom from a larger chiefdom.

Later studies by Underhill (1975) and Davis (1971) had trouble replicating Swanson's findings, partly due to unclear coding guidelines. However, Davis used a construct called 'societal complexity',

operationalized in four levels: (a) hunting and gathering, (b) simple horticultural, (c) advanced horticultural, and (d) agrarian. More recently, Peoples and Marlowe (2012) call these 'modes of subsistence', while Underhill (1975) calls this 'economic complexity'. Roes and Raymond (2003) take political complexity to directly reflect population size, where more levels supposedly indicate a larger population size, an assumption that has been criticized by others (Brown and Eff, 2010). Recently, Watts et al. (2015) based their measure of political complexity on the SCCS and EA: cultures were coded as low complexity when they had fewer than two levels of jurisdictional hierarchy and high complexity when they had at least two levels. Turchin et al. (2022) take a different approach for their measure of 'socio-political complexity' by aggregating 51 measures they think are representative of that concept (although there is no explicit motivation arguing why these measures should be considered relevant). In other cases, large-scale cooperation is seen as a feature for sustaining social complexity (Purzycki et al. 2022). Recent work by Purzycki and McKay (2023) and Lightner et al. (2023) equate the political complexity concept as used in the SCCS with social complexity. In sum, social/political complexity is a deeply ambiguous concept, resulting in opaque operationalizations.

In fact, this is a clear case of a 'jingle-jangle fallacy':, which refers to the false assumption that two measures capture the same construct because they bear the same name (jingle fallacy; e.g., the 51-variable operationalization of Turchin et al., 2022, of social-political complexity) – as well as the assumption that two measures with distinct names measure different constructs despite being very similar in nature. (jangle fallacy; e.g., social complexity and political complexity both being measured in terms of levels of jurisdictional hierarchies).

The most remarkable thing about the intellectual history of social/political complexity is that there has so far in the literature rarely ever been an explicitly theoretically motivated reason for using it as a concept. Evolutionary studies of religion rarely, if ever, refer to social complexity theory. Theoretical work on social complexity from Barton (2014), Kappeler (2019), and Stewart (1999) all agrees that social complexity is a deeply ambiguous concept and that may have its roots in Tylor's (1889) problematic theory of unilinear cultural evolution (Stewart, 1999). This theory is considered problematic because it argues that human societies follow a general, progressive evolutionary path from 'savage' irreligious groups to ultimately 'civilization'. The theoretical opacity of social complexity may have contributed to its inconsistent operationalizations, ranging from modes of subsistence to large aggregate measures of complexity. It appears that its continued use is more of a historical artefact rather than a theoretically motivated choice. In this sense, the social complexity construct is similar to the moralizing high gods construct. Therefore, as long as there is no clear theoretical motivation to include social complexity with an accompanying clear-cut operationalization with construct validity, it is perhaps best to give social complexity a similar treatment as the moralizing high

gods construct and stop including it in empirical research projects and theoretical frameworks. Instead, staying closer to cooperation is more fruitful for studying supernatural punishment, for the simple reason that evolutionary theories of religion are actually about cooperation (and not social complexity) at the causal-mechanistic level. An additional advantage is that cooperation has solid roots in evolutionary theory and game theory, whereas social complexity has none.

2.2.2.2. Prosociality, altruism, and cooperation

Prosociality, altruism, and cooperation are related concepts commonly used in both the biological and the social sciences, and are often even used interchangeably (West et al., 2012; Pfattheicher et al., 2022). Furthermore, only a quarter of studies that involve prosociality and altruism provide a definition (Pfattheicher et al., 2022). However, concepts such as cooperation and altruism have precise, formalized definitions that flow from evolutionary theory (West et al., 2012). These are distinct, and moreover, cooperation can be discerned into multiple different types of cooperation (e.g., Curry, 2016). Redefining these terms will only lead to conceptual confusion - an observation already made nearly 40 years ago by Dovidio (1984), still relevant today (Pfattheicher et al., 2022).

Cooperative behaviors can be classified into behaviors that confer either direct or indirect fitness benefits (West et al., 2012). Direct fitness benefits are yielded when the reproductive success of the one performing the cooperative behavior is increased. Indirect fitness refers to the components of fitness that are gained by aiding related individuals (also often called kin-selection). In other words, if one cooperates as to increase the fitness of related individuals without gaining direct fitness benefits themselves, one will gain indirect fitness. Therefore, by aiding closely related individuals at the dispense of oneself, an individual can still pass on their genes. Furthermore, cooperative behaviors can either be mutually beneficial or altruistic. They are mutually beneficial when the reproductive success of both the benefactor and beneficiary are increased (denoted as +/+). Cooperation is altruistic when the benefactor incurs damage to their reproductive success, while the beneficiary increases its reproductive success (-/+). (Large-scale) Cooperation constitutes an evolutionary problem because of free-riding. In the absence of social control or a punishment mechanism, free-riding would be the norm and lead to lower group-level fitness (Henrich et al., 2006). This is because free-riders benefit from the cooperative and altruistic behaviors of others, while not incurring any costs themselves. Over time, we would thus expect that cooperation would break down due to free riders increasing in relative numbers compared to cooperators. Cooperation can also be detrimental to both parties (denoted as -/-). This is the case with human punishment systems because they are costly to both the punitive party as well as the punished party at the individual level (Johnson, 2015).

Another important question pertains to what the meanings of ‘costs’ and ‘benefits’ are. According to Pfattheicher et al. (2022) the calculation of costs and benefits differ across research

traditions. Economic approaches, such as those using economic games (e.g., Lang et al., 2019; Laurin et al., 2012), typically involve conferring economic benefits to others at the expense of the actor. On the other hand, evolutionary approaches regard a behavior as costly if it reduces lifetime fitness (Pfattheicher et al., 2022). In other words, behaviors are only costly when they reduce the amount of offspring surviving to adulthood. As such, evolutionary and economic research traditions employ different usages of costs and benefits. This does not necessarily mean that one is 'better' than the other – it merely signals that scholars should exercise caution when framing their results in either economic or evolutionary terms. It does raise the question, however, how economic and evolutionary studies on the evolution of social behavior can be reconciled. This is particularly a challenge to accounts that seek to explain the evolution of cultural traits such as supernatural punishment purely on the basis of economic models of costs and benefits because it is often unclear if these costs and benefits reflect evolutionary fitness.

In connection to the morality-as-cooperation framework of Curry (2016) outlined earlier, it is important to recognize the diversity of prosocial behaviors when attempting to generalize one's study. This entails that one manifestation of prosociality (e.g., donations) does not automatically extend to other forms of prosociality (Pfattheicher et al., 2022). In other words, one particular instance of prosocial behavior may have positive effects on evolutionary fitness, while others are indifferent to it. Therefore, it is important to emphasize the breadth and diversity of the concept and to acknowledge that without this understanding we may obscure meaningful differences in the phenomena we wish to investigate (ibid.). A simple way to proceed, then, is to always provide definition of the concept. Subsequently, the operationalization should match this definition. Pfattheicher et al. (ibid.) illustrate this with the example that if a prosocial behavior is defined as intentional, the operationalization as well as the paradigm to evaluate it must reflect that intentionality.

Returning to how cooperation and prosocial behavior is regarded in discussions surrounding supernatural punishment, Bourrat and Viciana (2011) simply take cooperation to mean any social behavior susceptible to free-riding. In a more general review on the relationship between cultural evolutionary studies of religion and cooperation, Bendixen et al. (2021) state that 'when we discuss research on religion fostering prosociality or cooperation, we assume parochial (i.e., in-group) cooperation as opposed to universal or generalized prosociality'. Norenzayan et al. (2016) specifically focus on large-scale cooperation that is capable of driving the evolution of large-scale societies. As such, cooperation and prosociality are used in diverse ways in the literature. This is fine in principle; these concepts can be instantiated in different ways, after all. Crucially though, these instantiations ought to have a clear theoretical motivation behind them, and scholars should make clear how their

specific operationalization of cooperation/prosociality/altruism differs from other studies (especially when comparing results of different studies) and how this may have influenced the analysis.

3. Causal Analysis

Although causal claims are common in the evolutionary science of religion, only a few studies actually employ causal inference in a sound manner (e.g., Skoggard et al., 2020). This is problematic, as studies involving supernatural punishment often concern causal hypotheses. Broadly speaking, I have discerned three main hypotheses regarding the relationship between supernatural punishment and large-scale cooperation, each positing different causal structures. Firstly, the supernatural punishment hypothesis considers supernatural punishment to increase cooperation. Secondly, various predictions from the human behavioral ecology and the existential security hypothesis point to supernatural punishment developing as an evolved response to socio-ecological challenges, which includes resource scarcity, warfare, and also large-scale cooperation. Finally, the prosocial religions account put forward by Norenzayan et al. (2016) argues that supernatural punishment and large-scale cooperation co-evolve in a feedback loop. As such, the goal of this chapter is to delineate and substantiate the differences between these models, as well as to show how we might actually go about applying these models in research according to sound causal principles.

To proceed, I first introduce directed acyclic graphs before discussing basic causal inference theory. All of these graphs are drawn using dagitty.net (Textor et al., 2016). I discuss common pitfalls and biases that pervade the scientific literature and show how these may be dealt with. Second, I discuss how I apply causal analysis to the debate at hand by deriving causal models based on the supernatural punishment hypothesis, the prosocial religions framework, and the socio-ecological/existential security framework. Finally, I run data simulations from these models to show how confounding can harm statistical estimations.

3.1. Directed Acyclic Graphs

Directed Acyclic Graphs (DAGs) are graphical causal models that display the qualitative causal relationships between variables (Pearl, 1995). These relationships can take any functional form (e.g., exponential, linear, polynomial; Rohrer, 2016). An advantage of DAGs is that they require researchers to spell out their causal assumptions unambiguously. This is desirable, as almost all published literature on supernatural punishment only includes verbal models, resulting in underspecified causal structures. For example, the prosocial religions account of Norenzayan and associates (2016) argues that supernatural punishment is part of a larger religious 'package', yet what exactly constitutes a part

of this package is not fleshed out. Furthermore, they argue that large-scale cooperation can be reached via multiple cultural evolutionary paths, as well as that 'the causal effects of religious elements can interact with all of these domains and institution, and this causality can run in both directions' (6). However, these causal effects are not elucidated.

The relationships in DAGs represent a temporal sequence such that a cause necessarily precedes the effect. They are therefore acyclic because otherwise an effect would be able to causally influence their own antecedent preceding in time, which would lead to a grandfather paradox.

DAGs also distinguish between exposure and outcome variables. Exposure variables are also typically called predictor, treatment, or independent variables. Outcome variables are also usually called response or dependent variables. The causal flow between all variables originates in the variable(s) not caused by any other variables and follows the direction of the arrows.

DAGs are useful for three reasons: 1) They can be used to identify causal hypotheses from the theory, 2) they are intuitive visual representations that are transparent effective in communicating the assumptions drawn by researchers, and 3) they are useful for determining which variables to include or exclude in subsequent statistical analyses (McElreath, 2020).

3.1.1. DAGs and causal inference

Any valid kind causal inference depends on domain-specific assumptions (McElreath, 2020). Such assumptions are typically embedded within theories. Therefore, causal assumptions flow logically from a given theoretical framework. This also means that causal inference is not a mechanistic procedure; the steps required for causal inference depend on the assumptions drawn from the theory (Rohrer, 2016). This also entails that causal inference can never follow from just the data because any claims pertaining to causality are dependent on the assumptions that are drawn from the theory (e.g., X causes Y as a consequence of a hypothesis, not because of data). In other words, DAGs represent hypothesized causal relationships between variables.

One challenge for social science (and the 'softer' sciences in general) is that we basically never have access to a 'true' social reality and instead depend on proxy variables that we assume to represent that reality (Ember and Ember, 2009). This is because usually there is no direct way to manipulate 'true' social causes (Eronen and Bringmann, 2020). This brings up the entire discussion on whether and to what extent the proposed proxy variables are able to represent the 'true' social reality. As Rohrer (2016) observes, such proxy variables make it easier to establish clear causal relationships among variables but make it harder to generalize beyond that particular study. This challenge is particularly pervasive in large cross-cultural analyses of supernatural punishment due to the nature of how variables presumed to represent supernatural punishment and cooperation are encoded in highly varying ways (e.g., moralizing high gods to broad supernatural punishment). Therefore, whether 'true'

causal inference is possible for social sciences is debatable.

Another potential limitation is that feedback loops are hard to model using DAGs due to the aforementioned grandfather paradox. Yet, feedback loops are common in nature, ranging from biological and psychological processes to large-scale climate dynamics. I will illustrate this with a simple example. Intelligence is commonly thought to influence educational attainment. In turn, educational attainment could feed back into intelligence again, creating a feedback loop. Below I display an example of such a DAG (figure 1). The result of this DAG is that no effect can be estimated due to the temporal nature of how DAGs transmit causal influence. Put simply, causal influence is transmitted from intelligence to educational attainment such that intelligence precedes educational attainment temporally – therefore, the DAG below would actually predict that educational attainment is capable of influencing intelligence at a preceding point in time. Obviously, this is logically impossible.

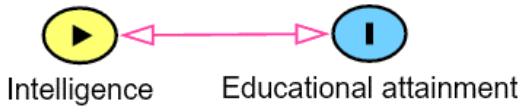

Figure 1. Here, intelligence predicts educational attainment. Educational attainment also predicts intelligence, but this is logically impossible due to educational attainment following intelligence chronologically.

Luckily, there are ways for DAGs to capture such prevalent feedback loops. One way of doing this is by adding temporal order to the DAG, such that 'childhood intelligence' is a different variable than 'adulthood intelligence', for example (figure 2). This allows us to estimate the influence of educational intelligence on adult intelligence, although other predictor-outcome relations may be explored as well depending on the research question and hypothesis.

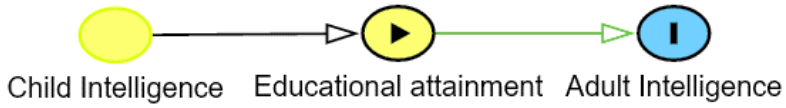

Figure 2. By making 'child intelligence' a different variable than 'adult intelligence', the grandfather paradox problem can be averted.

Another important point is that hypotheses and the theories they are derived from are underspecified. This entails that in many cases multiple causal structures are possible. As such, multiple DAGs are possible because of the hypotheses being vague or ambiguous due to their verbal nature (McElreath, 2020). Consequently, hypotheses do not imply unique causal models. Subsequently, the choice of statistical model also depends on those causal models. Therefore, statistical models may also reflect multiple causal models as well as multiple hypotheses. Thus, due to

limitations in prior knowledge, it is hard to adjudicate between DAGs and determine which is 'correct'. In such cases where multiple DAGs are possible for a hypothesis, it is useful to assess whether results differ meaningfully across analyses guided by different DAGs (Ellison, 2022).

3.1.2. Biases, confounders, and statistical conditioning

A common application for DAGs is to determine which variables to 'control for'. This is also commonly called conditioning, and it means to hold the value of a variable constant as to eliminate any potential influence it has on the target relationship (McElreath, 2020). DAGs show possible confounding relationships between the exposure variable and the outcome variable. Thus, by holding constant these confounders we are able to retrieve an unbiased estimate of the target relationship. However, it is a common practice and a held belief that adding more 'control' variables enhances the accuracy and validity of the study (e.g., the Turchin et al. 2022 paper). However, as Rohrer (2016, 28) notes, this is a 'methodological urban legend'. Simply adding as many variables as possible to a regression equation can lead to wrong conclusions due to a variety of biases, such as post-treatment bias and collider bias. I discuss these in-depth after first considering another common kind of bias, called confounder bias. I also discuss how these biases can be dealt with by utilizing specific statistical (and sometimes sampling) strategies.

One example of a confounding relationship is sometimes called a 'fork' (McElreath, 2020, 189 ff.). Forks are types of causal relationships whereby X and Y share a common cause Z (X ← Z → Y). The statistical consequence is that this generates a spurious correlation between X and Y if Z is not considered properly. For example, a correlation might arise between the rate of ice cream consumption and the number of sunburns due to a common cause: (hot) sunlight (see figure 3). Of course, ice cream consumption does not cause sunburns to arise, but sunlight causes people to consume more ice cream as well as get more sunburns. The standard way to deal with this problem is to hold the common cause at a constant value (McElreath, 2020). Holding sunlight constant will lead to the disappearance of the spurious correlation between ice cream consumption and sunburns (assuming there are no other confounding relationships). This is because variation in the common cause is what gave rise to the spurious relationship; eliminating that variation by holding the confounding variable constant removes the confounding effect.

If we imagine a scenario whereby there is an actual causal relationship between X and Y, confounded by Z, holding Z constant will not eliminate the relationship between X and Y. Controlling for Z merely eliminates the bias Z exerts on the relationship. Consider a study that attempts to investigate the influence of alcohol consumption on mortality rate. This relationship may be confounded by sex: men are more likely to drink more alcohol, as well as have a lower mortality rate

(see figure 4). Of course, there are dozens, if not hundreds of other possible confounding variables (age, ethnicity, exercise rate, and so on), but let us assume for the sake of illustration that sex is the only relevant confounding variable. If we were to study the effect of alcohol consumption on mortality rate without controlling for sex, this would lead to a biased result (partially dependent on the details of the sample, i.e., the proportion of men/women). However, if we control for sex, we can block the causal influence that flows from sex to mortality rate and alcohol consumption. Thus, controlling for sex will lead us to be able to estimate the direct, unbiased effect of alcohol consumption on mortality rate.

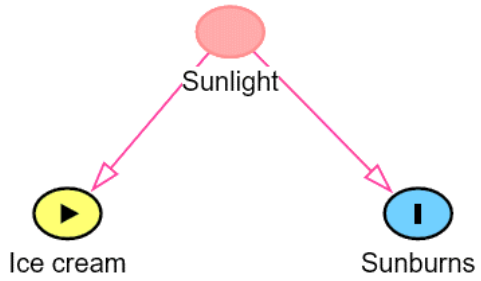

Figure 3. Sun light confounds the relationship between ice cream consumption and amount of sunburns, leading to a spurious association between ice cream consumption and sunburns.

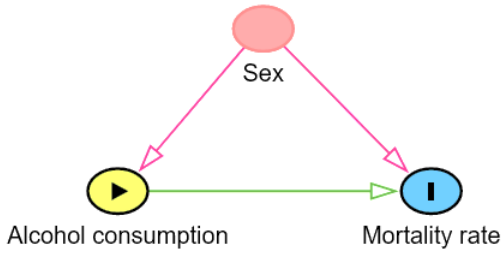

Figure 4. Sex confounds the relationship between alcohol consumption and mortality rate. To find out how alcohol consumption influences mortality rate, sex must be held constant.

Another type of causal relationship is called a mediator or a 'pipe' (McElreath, 2020, 189, ff.). Here Y is caused by Z, which in turn is caused by X (X → Z → Y). Conditioning on Z blocks the causal path from X to Y. Therefore, if we wish to know the direct relationship between X and Y, conditioning on Z should not be done because it eliminates the correlation. This is called posttreatment bias (Montgomery et al., 2018).

Consider the following example, taken from McElreath (2020). Suppose we wish to understand plant growth under different conditions of soil treatment against the development of fungus (see figure 6). Assumedly, the soil treatment influences fungus growth, and fungus growth hampers plant height. Plant heights are measured at an initial point in time before as well as after

having their soil treated. In this model, fungus is a mediator. However, fungus should not be included in the statistical model as a variable. This is because if fungus is included in the model, the causal path between treatment and plant height is blocked. The statistical result will show that the soil treatment had no effect on plant height, even though the opposite is actually the case. Another way of putting it is that once we know the fungus state (which is entirely the consequence of the soil treatment, the model assumed), additional information on the kind of soil treatment used bears no new knowledge on the plant height. This is because all relevant influence from the treatment is already contained within fungus. In sum, including variables in a statistical model without proper causal considerations may lead to posttreatment bias – a common occurrence in social science as shown by Montgomery et al., (2018), who found posttreatment conditioning in almost half of their sampled studies.

We might also consider a similar situation whereby X and Y are also directly causally related to each other (i.e., unmediated). In such situations, X exerts influence on Y both directly and indirectly (through Z). If we only wish to estimate the *direct* effect of X on Y we do need to hold Z constant. However, if the goal is to estimate the *total* effect X exerts on Y, we also need to consider the path that runs through Z; in this case, we should not condition on Z because this blocks that path. Consider, for example, the relationships between grades, self-esteem, and happiness (see figure 5). Achieving high grades may directly increase happiness for various reasons, such as increased fulfillment. Furthermore, grades will also influence the sense of self-esteem. Simultaneously, higher self-esteem also results in higher happiness. Therefore, self-esteem is considered a mediator in this causal model. The statistical modelling strategy we choose, that is, controlling for self-esteem or not, is dependent on the research question and hypothesis. If the goal is to estimate the *total* effect of grades on happiness, there is no need to control for self-esteem. If we were condition on self-esteem, we would block the causal flow from grades to happiness. However, if we only wish to determine how grades *directly* affect happiness, we do need to condition on self-esteem because this blocks that causal path.

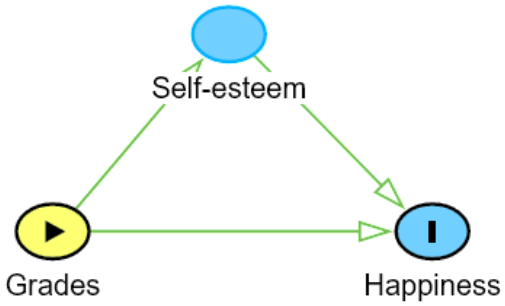

Figure 5. Self-esteem is a mediator in the relationship between grades and happiness, while there is also a direct influence grades transmits to happiness. If the research question involves ascertaining the direct effect grades have on happiness, self-esteem needs to be held constant. If the research question is concerned with the total effect grades transmits to happiness, self-esteem should not be controlled for.

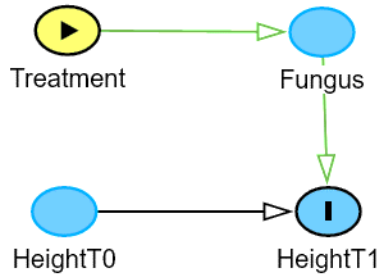

Figure 6. Fungus is a mediator between treatment and plant height. By including fungus in the statistical model, this blocks the path between the treatment and plant height. This would lead to the result that the treatment had no influence on the plant height. Therefore, fungus should not be included in the statistical model in such cases.

The final type of causal relationship I discuss is the collider. Here, X causes Z, and Y also causes Z (X → Z ← Y). This means that there is no causal relationship between X and Y. However, if we condition on Z we induce a spurious correlation between X and Y, called collider bias. In this sense, it is the opposite to a confounder. Consider the following example, where we are interested in understanding the relationship between acting talent and attractiveness (see figure 7). It often appears that actors are attractive. Attractive people tend to be cast more often in films and series. Additionally, we can assume that being a talented actor will lead to being cast in more films and series. However, there is no clear causal reason why attractiveness is related to acting talent. Yet, if we consider individuals who have a successful acting career, we will find a correlation between attractiveness and acting talent. In scientific research scenarios, collider bias is often the result of selection bias (McElreath, 2020). There may be correlations in the sample, but these correlations are the result of the way the sample is generated. Put simply, the spurious correlation is a feature of the sample, not of the population (where there might not be such correlation). Collider bias and posttreatment bias are the reasons why simply adding more variables for the purposes of statistical 'control' will lead to a biased result. Such procedures are often called garbage-can regressions or causal salad and should always be avoided (McElreath, 2020).

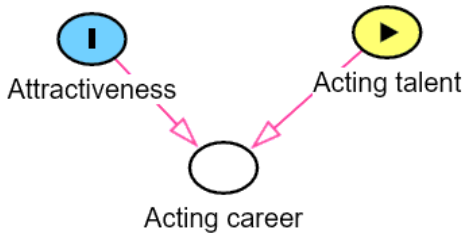

Figure 7. Controlling for 'acting career', a collider in the path between acting talent and attractiveness, induces a spurious association between acting talent and attractiveness.

These biases illustrate the importance of causal analysis in determining which variables to control or not to control for. There are numerous ways to statistically condition on a variable, once it has been determined that doing so will not lead to any biases or opening other confounded relationships.

The most common way in the social sciences is to add the variable that we wish to condition on to a multiple regression equation (Rohrer, 2016). In such equations, the outcome variable is then regressed on both the predictor variable and the confounding variable. This controls for the effect of the confounding variable, blocking the causal flow from it in the process.

Another way to condition on a variable is to perform a stratified analysis (Rohrer, 2016). This is commonly done with categorical variables, which take limited and fixed values, such as biological sex. Controlling for such variables entails stratifying them into their respective groups, e.g., men and women, effectively splitting the sample. If subsequent analyses show similar estimates for both groups, we could conclude that sex did not provide an alternative explanation.

Even though these are well-delineated ways to conduct causal inference, it must be noted that social and psychological reality is vastly more complex than such simple models are able to convey. Resultingly, this means that in certain complex constellations of variables, conditioning on one variable might reduce bias, but also increase bias if this variable is a collider on a different path (Rohrer, 2016). Furthermore, in such highly complex multidimensional models, it may also be unclear whether a variable is a collider, a mediator, or a confounder (ibid.). This distinction is hugely important, as this may influence the conclusions of the study. One way to deal with this is to compare different possible models and conduct multiple comparative analyses. Subsequently, these differences should be communicated for transparency. All this is to say is that we should exercise caution in interpreting causal results due the complex nature of messy social reality.

3.2. Deriving Causal Models

Multiple different theoretical models frameworks underly the literature discussed in the background section. As discussed in the introduction section and deepened in the background section, three main hypotheses can be distinguished in the literature. The first and most simple hypothesis is the supernatural punishment hypothesis, which states that supernatural punishment increases cooperation (the SPH model). The second hypothesis is that supernatural punishment evolves as a response to socio-ecological threats, including large-scale cooperation and instable climates (the socio-ecological model). The final hypothesis posits that supernatural punishment and large-scale cooperation co-evolve in a feedback loop (Norenzayan et al., 2016). Although all these hypotheses are

inspired by (cultural) evolutionary theories, they differ in terms of causal structures. Nevertheless, this does not imply that these hypotheses are necessarily exclusive to each other, in the sense that evidence for one of these hypotheses is evidence against of the other hypotheses. Crucially though, the hypotheses do differ in terms of predictor and outcome variables, which has implications for the DAGs and consequent data simulations. Below, I derive some potential DAGs associated with their relevant hypotheses.

3.2.1. The SPH model

The hypothesis that supernatural punishment is capable of stimulating large-scale cooperation has its roots in game theory (Johnson, 2005). The reason why large-scale cooperation is seen as such a fundamental issue in evolutionary theory is that cheating would be the default choice of action in such cooperative situations (Powers et al., 2021). This is because the potential cost of cooperation is higher than the potential cost of cheating being punished. Therefore, the ubiquity of large-scale cooperation in human societies presents a theoretical puzzle to the evolutionary social sciences (Henrich et al., 2016). The supernatural punishment hypothesis presents a solution to this challenge by arguing that supernatural punishment alters the payoff structure in large-scale cooperative games such that the cost of cooperation is lower than the cost of being punished. However, Johnson (2005; 2015) and Schloss and Murray (2011) point out that the SPH primarily focusses on punishment avoidance in small-scale societies. Furthermore, the SPH as formulated by Johnson (2015) emphasizes individual selection over group selection. Of course, this approach does not exclude the possibility of supernatural punishment evolving via group selection, nor does it exclude that supernatural punishment acts beyond the level of small-scale societies. This is a fact that Johnson (2015) also recognizes and supports. In fact, Johnson argues that such accounts can be subsumed under his hypothesis, as they are logically consistent with each other. Additionally, Johnson and Krüger (2004, 160) state that supernatural punishment can also explain cooperation in large-scale groups, whereas standard theories of reciprocal altruism and kin selection cannot: 'None of these theories solve the puzzle of why humans continue to cooperate in large groups of genetically unrelated strangers, in single-shot interactions and when gains from reputation are negligible—that is, where all of those mechanisms are inapplicable'. Thus, as this thesis mainly focusses on the puzzle of large-scale cooperation, I will frame Johnson's hypothesis as such.

The SPH model translates to a simple DAG with four variables. Beside supernatural punishment as a causal influence on large-scale cooperation, Johnson (2005; 2015) also emphasizes that other forms of human punishment, fueled by entrenched norms and laws, can enhance cooperation. Furthermore, Johnson (ibid.) argues that supernatural reward plays a role in facilitating large-scale cooperation, although he emphasizes that any potential effects of punishment will be

stronger than reward, based on common findings in social science and psychological studies that show that the 'stick' (punishment) is more powerful in steering behavior than the 'carrot' (reward). Therefore, a potential DAG of Johnson's supernatural punishment model might look as follows:

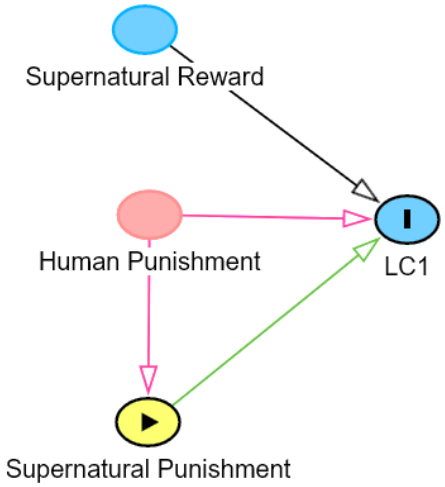

Figure 8. Supernatural punishment causes large-scale cooperation, with human punishment as a confounder. Supernatural reward also causes cooperation.

Here, 'LC1' refers to large-scale cooperation. The model draws the following causal assumptions:

LC1 ← supernatural punishment + human punishment + supernatural reward

The model argues that larger-scale cooperation is caused by both human punishment and supernatural reward in addition to supernatural punishment. Human punishment contributes to large-scale cooperation because the punishment mechanism may stabilize cooperative behaviors (LC1 ← human punishment). Johnson attributes a stronger role to supernatural punishment than human punishment because human punishment is fallible to second-order free-riding while supernatural punishment is not (LC1 ← supernatural punishment). A similar role is attributed to supernatural reward because of the 'carrot' being a less powerful mechanism for stimulating cooperative behavior than the 'stick' (LC1 ← supernatural reward).

Supernatural punishment ← human punishment

This model posits that supernatural punishment arises as a consequence of human punishment. This relationship can come about in at least two ways. The first is that the presence or absence of human punishment creates a potential selection pressure on the development of supernatural punishment (Johnson, 2015). For example, in a society with very little human punishment, yet with many large-

scale cooperative interactions, the development of supernatural punishment would be adaptive. The second way is that successful secular punitive institutions, such as a police forces and legal systems, potentially make supernatural punishment obsolete (Norenzayan, 2013).

To estimate the influence of supernatural punishment on large-scale cooperation, confounding relationships need to be held constant. In this model, human punishment is the only confounder. Additionally, supernatural reward could also be held constant because it also transmits an influence on large-scale cooperation, but this is optional.

3.2.2. The HBE model

The basic idea of socio-ecological models stemming from human behavioral ecology (HBE) is that some human behaviors and cultural traits, such as supernatural punishment, may be naturally selected adaptations to socio-ecological challenges, such as warfare, natural disasters, and large-scale cooperation (Johnson, 2015; Purzycki et al., 2022). Subsequently, gods grow to evolve as a response to these threats by virtue of them altering the pay-off structures of cooperative games such that people would be more inclined to cooperate, leading to an evolutionary advantage (Purzycki et al., 2022). Therefore, HBE models are not incongruent with the hypothesis that supernatural punishment can increase large-scale cooperation. However, HBE typically emphasizes the evolutionary adaptiveness and variation of human behavior in response to environmental conditions. This means that such studies would in our case focus on supernatural punishment as an outcome of socio-ecological conditions.

The HBE model of supernatural punishment also draws assumptions derived from the existential security hypothesis. This hypothesis states that increased prevalence of existential threats increases religious practices and belief (e.g., Baimel et al., 2021). Such existential threats include, but are not limited to, ecological stress (drought, natural disasters), resource scarcity (lack of water), agricultural productivity, and war (Botero et al., 2014; Skoggard et al., 2020; Turchin et al., 2022). However, such existential threats do not only increase religious beliefs and practices, some of them may also influence cooperation. For example, warfare may increase both beliefs in supernatural punishment as well as demand more intra-group cooperation (Turchin et al., 2022). An increase in agricultural production may lead to new cooperative dilemmas, as food distribution increasingly becomes a task that requires cooperation at increasingly larger scales (Baumard et al., 2015). Finally, group size is inhibited (or perhaps favored) by existential threats such as warfare and natural disasters. Furthermore, without large-scale cooperation, larger groups are more likely to fission or experience internal conflict due to free-riding (Norenzayan et al., 2016).

Taken together, the hypothesis that supernatural punishment develops as an evolved

response to large-scale cooperation results in the following possible DAG, based on the HBE framework and the existential security hypothesis:

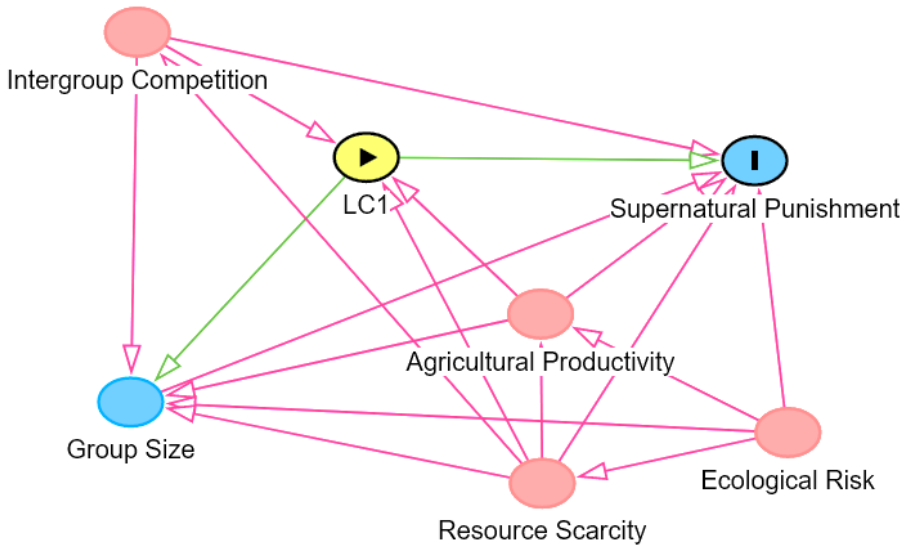

Figure 9. The HBE model considers supernatural punishment as an evolved response to socio-ecological selection pressures, such as agricultural productivity, ecological stress, resource scarcity, intergroup competition, and large-scale cooperation.

Here, 'LC1' refers to large-scale cooperation. Following the theoretical model, the following causal assumptions are drawn:

> *SP ← Intergroup competition + LC1 + agricultural productivity + group size + resource scarcity + ecological risk*

Ecological risk, resource scarcity, agricultural productivity, intergroup competition, group size and large-scale cooperation are all assumed to have a causal relationship with supernatural punishment. Existential security models predict that that beliefs in supernatural punishment increase during times of duress, meaning that conditions of e.g., war (SP ← Intergroup competition; Henrich et al., 2019), lack of water (SP ← Resource scarcity; Skoggard et al., 2020), and natural disasters (SP ← Ecological risk; Botero et al., 2014) all contribute to increased supernatural punishment beliefs. Furthermore, agricultural productivity is related to supernatural punishment because stable food production led to increased material and existential security (SP ← Agricultural productivity; Peoples and Marlowe, 2012). Larger, successful groups are both more likely to spread than smaller-scale groups, which simultaneously increases the spread of supernatural punishment beliefs (SP ← Group size;

Norenzayan et al., 2016). Finally, large-scale cooperation is considered to causally influence the presence of supernatural punishment beliefs because large-scale cooperative situations create a selection pressure on the development of such beliefs to manage and curtail possible free-riding (SP ← LC1; Purzycki et al., 2022).

LC1 ← Intergroup competition + agricultural productivity + resource scarcity

This model posits that large-scale cooperation develops as a product of the combined influence of intergroup competition, agricultural productivity, resource scarcity. Intergroup competition, such as warfare, necessitates and increases large-scale cooperation due to increased contacts between unrelated individuals (LC1 ← Intergroup competition; Turchin et al., 2022). Furthermore, agricultural productivity increases large-scale cooperation because increased food production will lead to novel large-scale cooperative dilemmas regarding e.g., food distribution, meaning that societies that have been able to sustain highly productive agricultural systems are thus more likely to feature large-scale cooperation (LC1 ← Agricultural productivity; Peoples and Marlowe, 2012). Resource stress increases large-scale cooperation, as argued by Skoggard et al., (2020) and Ember et al. (2018), because resource sharing in times of scarcity is considered a way to build social capital (LC1 ← Resource scarcity).

Agricultural productivity ← resource scarcity + ecological risk

This model posits that agricultural productivity develops as a response to ecological circumstances as well as resource presence. For instance, in locations where droughts or instable climates are prevalent, agricultural productivity will be low, whereas locations that feature stable climates with steady rainfall are more optimal to facilitate agricultural development (Agricultural productivity ← ecological risk; Botero et al., 2014). Additionally, locations where natural resources such as water are scarce, agricultural productivity will likely be low as well (Agricultural productivity ← resource scarcity; ibid.).

Group size ← Intergroup competition + LC1 + Agricultural productivity + Resource scarcity + Ecological risk

Large-scale cooperation influences group size, as more cooperative groups are more likely to expand (Group size ← LC1; Norenzayan et al., 2016). Furthermore, intergroup competition influences group size, as during e.g., warfare, group size will become smaller due to casualties (Group size ← Intergroup competition). In addition, increased food production by virtue of increased agricultural productivity will facilitate population growth (Group size ← Agricultural Productivity). Resource scarcity and ecological risk may inhibit group size by virtue of e.g., water scarcity (Group size ← Resource scarcity; Snarey, 1996) and prevalence of natural disasters (Group size ← Ecological risk; Botero et al., 2014).

Resource scarcity ← Ecological risk

In this model, the abundance or scarcity of resources is influenced by ecological threats. Long periods of draught may influence the scarcity of water, for example (Botero et al., 2014).

There are two ways to estimate the influence (either total or direct; they are equivalent in this model) of large-scale cooperation on supernatural punishment. One way is to hold constant the variables agricultural productivity, ecological risk, resource scarcity, and intergroup competition. The second way is to hold constant the variables agricultural productivity, group size, and intergroup competition.

3.2.3. The prosocial religions model

The prosocial religions account of Norenzayan et al. (2016) presents supernatural punishment and large-scale cooperation as co-evolving, and it has firm roots in cultural evolutionary theory. Although Norenzayan and associates primarily discuss Big Gods and other elements of the religious 'package' (which includes rituals and costly displays as well), the actual mechanism in question responsible for engendering large-scale cooperation is supernatural punishment (Johnson, 2015). The prosocial religions model argues that supernatural punishment spread because of their contribution to large-scale groups. In turn, large-scale societies were more likely to transmit supernatural punishment beliefs. They further argue that other religious elements, such as rituals and signs of devotion to a punitive god assimilated into a package of religious beliefs and practices, which stimulate cooperation among co-religionists. Additionally, their framework argues that such cultural groups entertain a relative cultural evolutionary advantage in cases of intergroup competition over resources. Ultimately, cultural group selection explains how such large-scale groups featuring prosocial religious elements evolved based on the competitive advantages they bestow.

As such, a possible DAG for this model looks as follows:

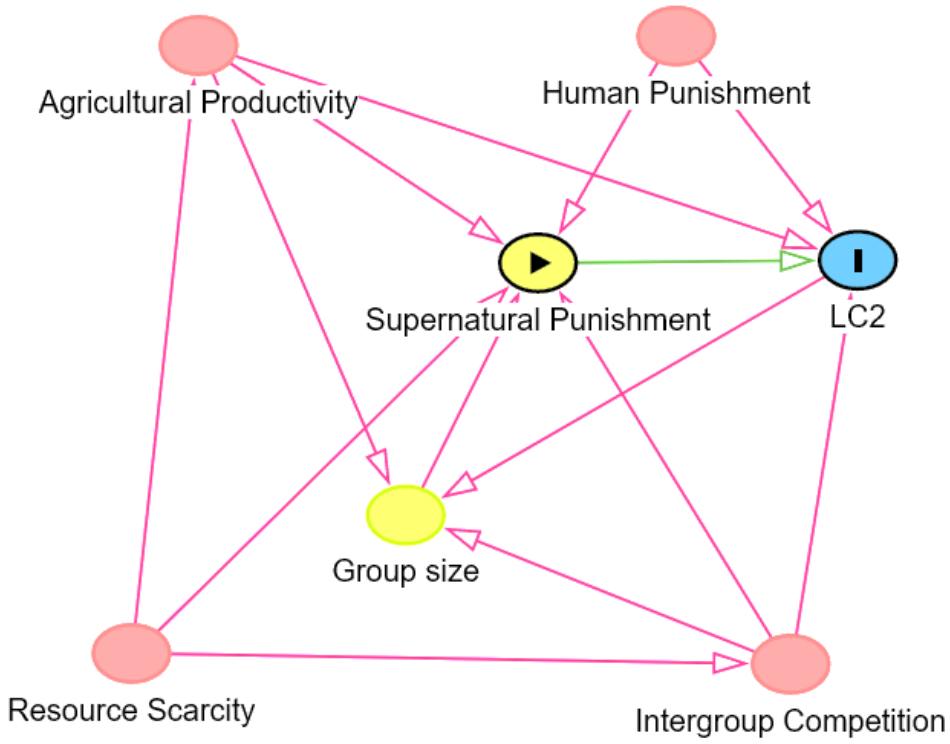

Figure 10. This causal model of the prosocial religions account is untestable because the influence of large-scale cooperation on group size feeds back into supernatural punishment.

'LC2' represents large-scale cooperation in complex settled societies, which is similar to LC1 but distinct in a few important ways. Norenzayan et al. (2016) are interested in explaining the rise of stable large-scale societies. As Schloss and Murray (2011) and Johnson (2015) point out, the prosocial religions account is primarily associated with cooperation enhancement, whereas Johnson's supernatural punishment hypothesis is primarily associated with punishment avoidance. As such, the prosocial religions account seeks to explain how large-scale cooperation is *maintained* in large societies. Furthermore, although both theoretical accounts emphasize supernatural punishment as a key mechanism, Norenzayan et al. (2016) lean more on Big Gods, whereas Johnson is more focused on spirits. This model draws the following causal assumptions:

> *LC2 ← Supernatural punishment + human punishment + intergroup competition + agricultural productivity*

This model considers supernatural punishment as the predictor and large-scale cooperation in complex societies as the outcome (LC2 ← SP). Norenzayan et al. (2016) argue that cultural evolution selected for a suite of religious beliefs, including supernatural punishment, as a response to intergroup competition over resources in settled societies. This, in turn, contributed to increases in cooperation

(LC2 ← intergroup competition). Furthermore, the prosocial religions account argues that powerful secular institutions that increased trust and solidarity (referred to as human punishment here) reduced the selective forces that selected for supernatural punishment and simultaneously increased large-scale cooperation (LC2 ← Human punishment). Finally, the prosocial religions account puts significant emphasis on agricultural productivity, as it seen as a key prerequisite for the mechanistic interaction between supernatural punishment and large-scale cooperation because it made settled, complex societies economically possible in the first place (LC2 ← Agricultural productivity).

SP ← Resource scarcity + agricultural productivity + group size + intergroup competition + human punishment

The prosocial religions model argues that beliefs in supernatural punishment spread by virtue of intergroup competition over scarce resources (SP ← Intergroup competition + Resource scarcity). Furthermore, increasingly larger-scale societies also increasingly contributed to the transmission of beliefs, entailing that supernatural punishment increases as a consequence of group size (SP ← Group size) . Additionally, the presence of successful secular punishment institutions modifies the selection pressures on the cultural evolutionary development and spread of supernatural punishment (SP ← Human punishment). Finally, agricultural productivity is linked to supernatural punishment due to the adoption of different life-history strategies. That is, food abundance may lead to slower life-history strategies and therefore increase beliefs in supernatural punishment (SP ← Agricultural productivity; Baumard et al., 2015). Finally, resource scarcity may lead to reduced existential security and therefore increased belief in supernatural punishment (SP ← Resource scarcity; Botero et al., 2014).

Group size ← Agricultural productivity + LC2 + intergroup competition

The model posits that group size is the result of the combined forces of agricultural productivity, large-scale cooperation, and intergroup competition. The development of agricultural productivity made large groups feasible in the first place because without stable food supply, group size will dwindle (Group size ← Agricultural productivity). Furthermore, intergroup competition influences group size either through assimilation (whereby groups merge together), through warfare (whereby bloodshed reduces group size), or through cultural group selection (whereby more cooperative groups are more successful in in intergroup competition; Group Size ← Intergroup competition). Finally, Norenzayan et al. (2016) argue that complex societies will fission or become unstable without widespread large-scale cooperation (Group size ← LC2).

Intergroup competition ← Resource scarcity

The prosocial religions account does not give an extensive description of the conditions under which intergroup competition occurs. It is only mentioned that intergroup competition over scarce resources is capable of driving large-scale cooperation (Intergroup competition ← Resource scarcity).

Following the rules of causal inference, the effect of supernatural punishment on large-scale cooperation in complex societies cannot be estimated. This is due to the causal chain Supernatural punishment → LC2 → Group size → Supernatural punishment. Supernatural punishment causes itself due to the feedback effect running from large-scale cooperation to group size and back into supernatural punishment, which is causally invalid. Therefore, there is no way to use the prosocial religions model to estimate the direct effect of supernatural punishment on large-scale cooperation.

Several alternatives are possible though. One way would be to model the feedback effects by measuring the same variable at different moments in time, as illustrated below.

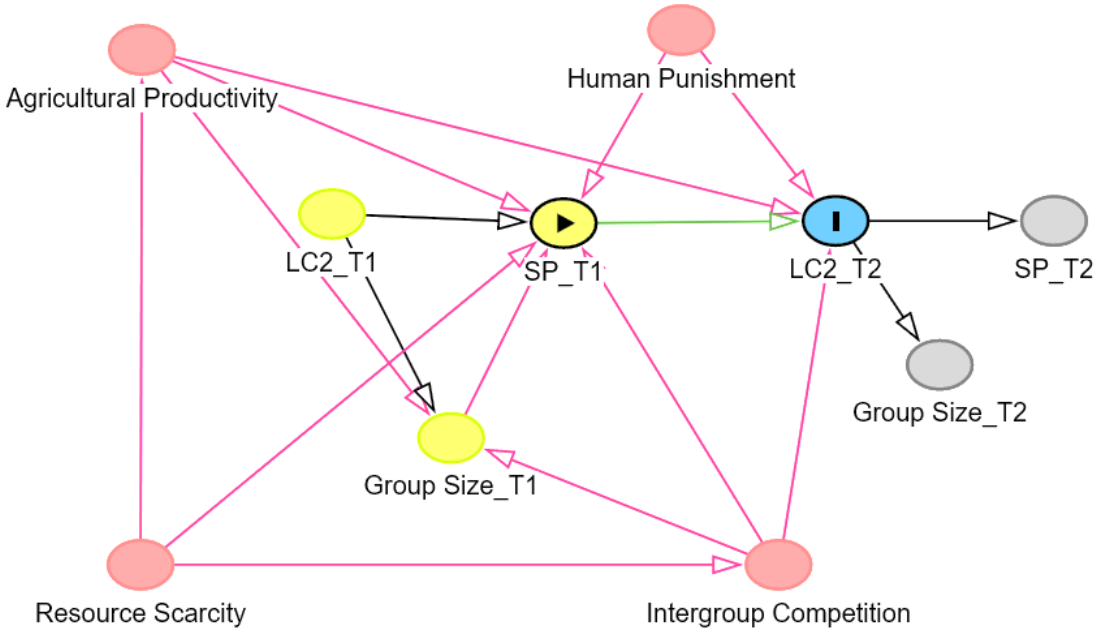

Figure 11. The co-evolutionary nature of the prosocial religions model can be accounted for by including multiple time points for each variable, as is shown here with SP2 and group size 2.

Such a model would arguably stay closer to the co-evolving nature of the prosocial religions framework, as it acknowledges the feedback mechanisms theorized in the model. However, it would also make an empirical test of the model impractical, as it would require measurements of the variables at two different points in time. Reliable large-scale cross-cultural data on this is currently unavailable (although see Turchin et al., 2022).

Another alternative is to consider the alternative causal direction that runs from large-scale

cooperation to supernatural punishment (see figure 12). Of course, this results in answering a different research question, but it would still be a test of the prosocial religions model. Thus, if we wish to estimate the influence of LC2→Supernatural punishment, we will need to hold constant Agricultural productivity, Human punishment, Group size, and Intergroup competition.

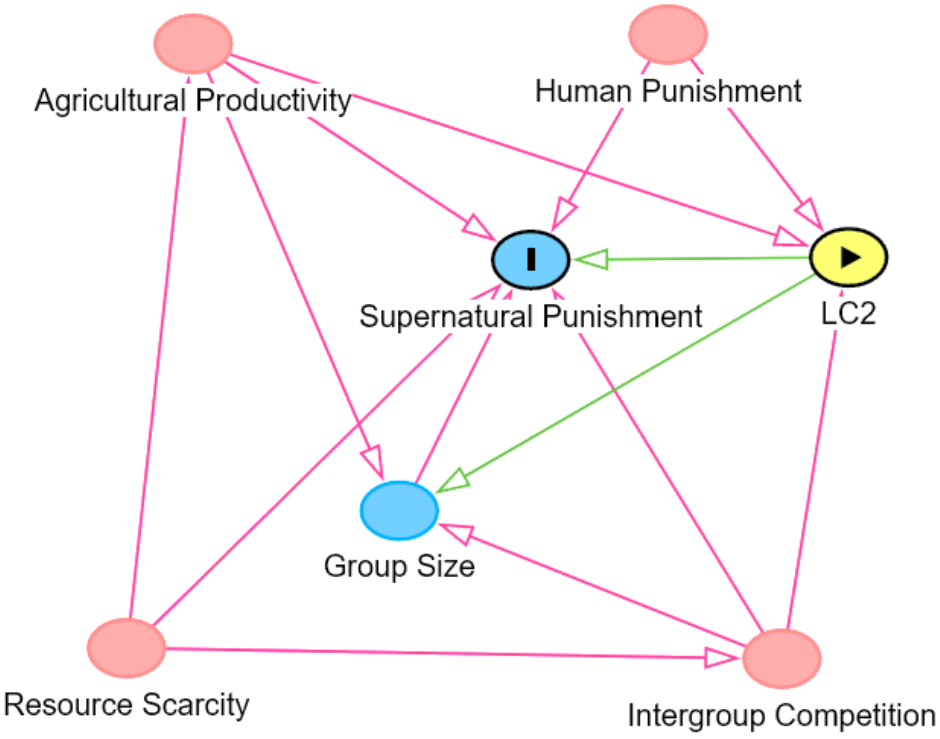

Figure 12. The reversed prosocial religions model, whereby LC2 causes SP instead of the other way around, is testable as it is no longer the case that a variable causes itself.

3.3. Simulation

Scientific studies often follow a specific process for testing causal hypotheses. A typical scientific workflow (McElreath, 2020) starts with specifying a theoretical estimand, which represents the relationship that is to be investigated. Subsequently, a DAG is drawn that represents the logical connections based on the theoretical model. The third step is to simulate data from this DAG. This entails that random data is generated following specified rules. This allows researchers to explore how factors that could confound the study may influence the results. The subsequent steps include designing a statistical model to actually estimate the effect on real data.

In this thesis, I focus only on the first three steps. The simulations aim to determine whether predefined estimates can be recovered. If the simulation is able to retrieve the predefined estimate, it means that the target relationship can be empirically tested, provided that confounding factors are

appropriately addressed and the model's logic is valid. If the predefined estimate cannot be recovered, it suggests that an empirical test will always yield biased results, which would lead to flawed inferences.

The causal models are simulated in the following way. Simple random normal distributions for the variables are used. These normal distributions are assumed to be centered around a mean value of 0 and have a standard deviation of 1. Furthermore, the variables are also programmed to have some measurement error included for realism. To estimate the effect of exposure variables to outcome variables, the effect of supernatural punishment on large-scale cooperation (or the other way around) is assumed to be 0.5 by the model. This is the predefined estimate that we are attempting to retrieve. If it is unable to be retrieved, this is a sign of the model being biased. The practical meaning of the 0.5 value is that every time the independent variable (e.g., supernatural punishment) changes by one standard deviation, the estimated outcome variable changes by 0.5 on average.

Using the simulated data, linear regression models are ran whereby the outcome variable is regressed on the predictor variable as well as on the predictor variables combined with the other confounding variables. Subsequently, these regressions are replicated 1000 times and then averaged. Wherever possible, each graph includes both a controlled and a confounded model. The results of the simulation of the models are displayed in probability distribution plots.

For simplicity's sake, all other relationships are considered to have an effect of 0.4. It is certain that such values do not reflect the true relationships, but the problem is that the theory is underspecified, which makes it difficult to establish values to represent the relationships among variables. The goal here is to illustrate 1) how predefined estimates can be recovered by accounting for confounders, and 2) how confounders lead to biased results if unaccounted for.

Below, I show 4 graphs (one for each causal model), each of which displays both a confounded and a controlled model.

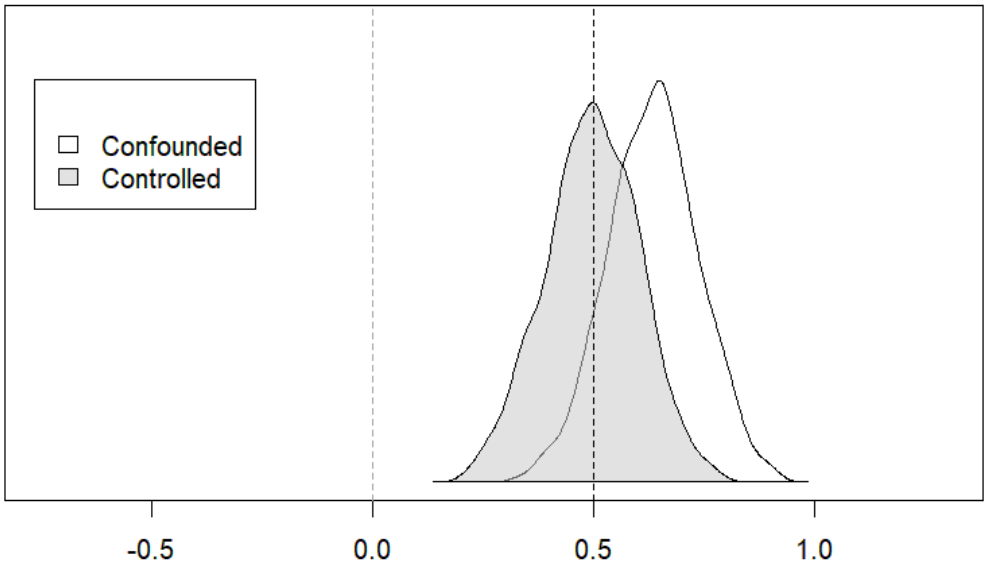

Figure 13. SPH model. When human punishment is unaccounted for, the predefined estimate of 0.5 is not retrieved, showing the impact of confounding. When holding human punishment constant, we retrieve the predefined effect of 0.5. When we merely regress large-scale cooperation on supernatural punishment, we do not retrieve this effect, and instead retrieve a biased effect of 0.7.

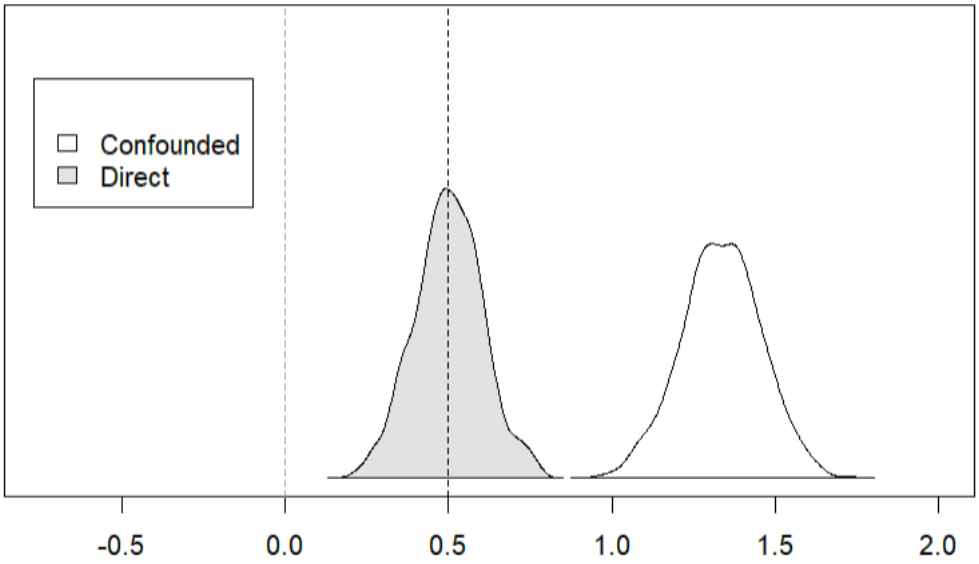

Figure 14. HBE model. The simulations of the HBE model show the possible effects confounding has on the estimate. When holding constant all possible confounders we retrieve the predefined effect of 0.5. The uncontrolled regression model, where supernatural punishment is solely estimated through the effects of large-scale cooperation, reveals an almost tripled effect.

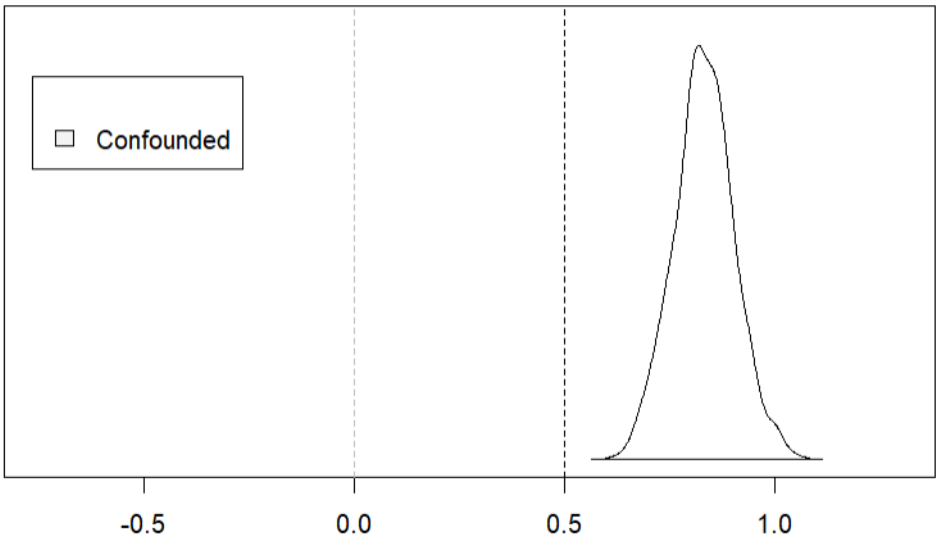

Figure 15. Prosocial religions model. Due to Supernatural punishment feeding back into itself via the pathway through large-scale cooperation → Group size → Supernatural punishment, it is impossible to retrieve the predefined estimate of 0.5, meaning it is impossible to retrieve an unbiased result.

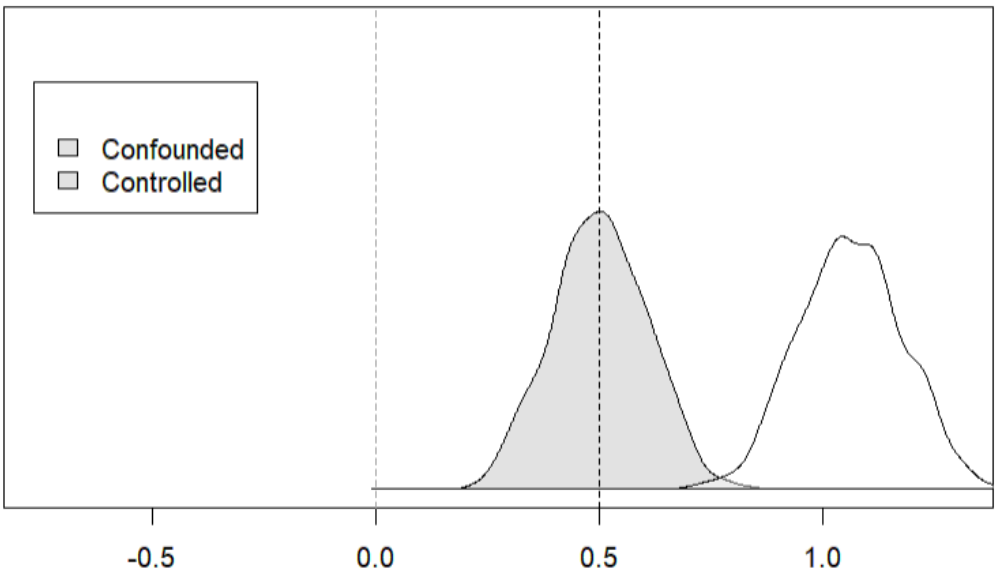

Figure 16. Reversed prosocial religions model. The LC2→SP association can be successfully estimated if confounding variables are held constant in contrast to the SP→LC2 association.

4. Methodological issues

Evolutionary studies of supernatural punishment and cooperation face pervasive challenges in data collection, validity, cross-cultural database design, and causal inference. This chapter first provides an in-depth discussion of each of these issues before moving on to how such challenges may be addressed in the second section.

4.1. Methodological Challenges

Investigating supernatural punishment evokes myriad methodological challenges. For example, many studies rely on cross-cultural databases, but the quality of the inferences made from these is heavily reliant on the data quality of the respective ethnographic sources (Ember and Ember, 2009; Lightner et al., 2023). Furthermore, such data faces issues of validity, as it is unclear whether the coded data is representative of a cross-cultural reality (Ember and Ember, 2009). In the same vein, the design choices that go into the development of cross-cultural databases has consequences for the variables recorded (e.g., whether the database has a variable for moralizing *high* god, but not for moralizing god) in it. This constrains the kinds of research questions can be operationalized, as there is a strong possibility of not all theoretically required variables being recorded. Moreover, all cross-cultural comparative research must grapple with Galton's problem, which is the statistical challenge that cultures are not non-independent statistical points due to cultural diffusion caused by shared history and cultural contact (Naroll, 1965; Ember and Ember, 2009). In other words, cultures are auto-correlated based on cultural diffusion and shared cultural history. Finally, the field has had recent controversies concerning causal and statistical inference on the basis of ethnographic records (Turchin et al., 2022; Purzycki et al., 2022), highlighting the importance for a more rigorous causal approach to the investigation of supernatural punishment. I review each of these issues in turn.

4.1.1. Ethnographic data source quality

The most common manner in the literature of assessing cross-cultural patterns of human variation is to use the SCCS or the EA (Purzycki and McKay, 2023). These cross-cultural databases contain quantitative information about cultures, translated from ethnographic research, missionary reports, and travelers' accounts. These large-scale cross-cultural databases converted that information into analyzable data. However, recently, worries have been raised about the quality of such data (Lightner et al., 2023; Purzycki and McKay, 2023; Watts et al., 2022). This is because these databases fully rely on previously written ethnographies and missionary reports that are sometimes over a century old (Ember and Ember, 2009). Ethnographic standards were very different a century ago as compared to the present day. These older ethnographies suffer from poor methodological standards and are rife with ethnocentrism, racism and other biases (Purzycki and McKay, 2023; Lightner et al., 2023).

An example of such a bias common in 20th century Christian missionaries is the 'primitive monotheism' bias (Watts et al., 2022). This was the belief that cultures all start with an Abrahamic-like god but were corrupted by polytheistic beliefs. While this may seem obviously wrong to present-day researchers, even trained anthropologists and ethnographers are not immune to bias. This is because their research is often guided by frameworks and corresponding theoretical assumptions. One such framework popular in the 20th century was the theory unilateral socio-cultural evolution,

which argued that cultural groups evolve culturally from primitive savages to civilized societies with Eurocentric norms and values (Tylor, 1889). Therefore, it is incredibly important to be aware how the data in cross-cultural databases came to be as the quality of the data often leaves much to be desired (Purzycki and McKay, 2023; Lightner et al., 2023).

Despite that, cross-cultural datasets rarely discuss data quality, yet the data quality is foundational to the reliability and theoretical validity of such datasets (Ember and Ember, 2009). Those using the cross-cultural database may not be aware of systemic bias present in the dataset, as has been the case for the last sixty or so years before scholars realized that deep problems are associated with cross-cultural database relying on old ethnographic and missionary accounts. Now that awareness over these biases has grown, scholars that continue to use database such as the SCCS and EA should be transparent over how the quality of the data may skew the results of their study as well as model the data generation process (Watts et al., 2022). A way to address this could be to give each source a data quality score and then give more weight to higher scored sources in the analysis-phase (ibid.).

Theories and models are built on data. Yet, many contemporary evolutionary theories of religion are shaped by such flawed data. As Freedman notes, 'good models are hard to build on the basis of bad data' (1985, 345). To properly understand the relationship between supernatural punishment and large-scale cooperation, high quality data are foundational to this cause. It is therefore important to exercise caution in interpreting the results of cross-cultural database studies that rely on biased ethnographic records.

4.1.2. Validity and ethnographic data

Establishing construct validity is of central importance for successful research design. This involves determining whether a specific measure is an accurate representation or proxy of what it is supposed to measure (Ember and Ember, 2009). This is especially challenging for studies sampling from pre-coded data from cross-cultural database. Luckily, there have still been established procedures that allow us to grasp the validity of cross-cultural measures. Ember and Ember (2009) argue that one such procedure begins with a theoretical definition of a variable of interest; this may be done verbally and/or mathematically. Subsequently, this theoretical construct needs to be operationalized. This can be done by devising a scale and spelling out when a certain case falls on such a scale.

In cross-cultural research on supernatural punishment and large-scale cooperation, there are not many unequivocally accepted and theoretically standardized measures available. In such cases, other ways of establishing validity, such as content validity can be pursued. Content validity revolves around to which degree 'a specified domain of content is sampled' (Nunnally, 1978, 91). When

developing measures for abstract theoretical constructs, such as reputational status or social complexity, this may lead one to measuring many variables. This is the operational strategy used by Turchin et al. (2022), who use a 51-variable aggregate measure for social complexity and a 7-variable aggregate for moralizing supernatural punishment. The idea is that such a measure is more likely to be valid if it taps multiple relevant domains (Ember and Ember, 2009). Of course, this requires that these variables actually tap the relevant dimensions of variation. This is particularly problematic when a concept is not sufficiently theoretically defined, as is the case with social complexity.

Such challenges are a common obstacle in empirical social scientific research. The main problem is that there is no direct access to the 'true' social reality (Eronen and Bringmann, 2020). We are only able to approximate it via proxy measures that supposedly represent that reality. Using those proxy variables it may become easier to establish clear causal relationships between variables (internal validity), but it may hinder generalizing beyond the study (external validity; ibid.). This may seem like an unsurmountable challenge, but one way this can be addressed is to explicitly address operational differences between different empirical research projects (Bringmann et al., 2022).

Another way to establish validity is by showing that a measure has a high correlation with generally accepted measure. This assumes that the correlation is caused by these measures measuring the same construct (and not that the correlation is incidentally caused by other variables). Studies that use differing scales that measure the same theoretical constructs can estimate the intercorrelation of the scales to establish convergent validity. For example, Lightner and others (2023) use this approach to show how diverging social complexity scales intercorrelate.

4.1.3. Cross-cultural database design

Cross-cultural databases provide crucial infrastructure for cross-cultural research. However, their design architecture can pose limitations, such as permitting what variables can be recorded in the first place. For example, some databases, such as the SCCS, only return a positive value for 'high god' if an ethnographer recorded a culture as having 'a spiritual being who is believed to have created all reality and/or to be its ultimate governor, even though his/her sole act was to create other spirits who, in turn, created or control the natural world' (Murdock and White, 2006). However, Lightner et al. (2023) argue that the presence of moralizing gods in small-scale societies is underestimated by using such variables. This is in part by virtue of the coding scheme, whereby moralizing gods and moralizing *high* gods are conflated into a single category. This manner of coding results in many false negatives, as small-scale societies are recorded as having no moralizing religion (due to the lack of a creator deity), even though they do (ibid.). However, there is no known mechanism of how a *creator* deity is supposed to influence social complexity as opposed to any other moralizing deity. This shows how

methodological considerations, namely the design of the cross-cultural database, constrain both the theorization and conceptualization of the key variables. Therefore, the design considerations in the development process of cross-cultural databases are pivotal in determining and limiting the research questions we can pose.

Recently, concrete suggestions have been made to overcome some of these challenges in constructing cross-cultural databases from ethnographic records (Watts et al., 2022). Watts and his colleagues argue that database creators should be transparent about how variables are defined, how coding decisions have been made, and how limitations at the ethnographic level are addressed. They offer several solutions. For example, an alternative to coding at the level of the cultural group is to code at the level of the ethnographic source level. This avoids the possibly pernicious situation where different ethnographic sources have to be aggregated, yet which claim diverging things about their respective culture. Furthermore, coding at the ethnographic levels allows for different time and place foci. To chart potential ethnographic biases, it is also useful to code meta-data. For example, one might want to exclude missionary reports if there is reason to suspect that these are heavily biased. Such solutions provide directions in dealing with the challenges associated with databases, such as the SCCS, that have unclear variable definitions and dubious coding decisions.

Similarly, Slingerland et al. (2020) provide recommendations for designing cross-cultural databases. In their view, the most difficult problem for building cultural databases is defining units of analysis and translating thick qualitative data to quantitative data. They recommend that the cultural units should be carefully chosen to ensure that they are directly comparable and have a specific time and place focus. A database should have a clear theoretical goal, but this should not be too narrow due to limited usefulness to the broader research community (ibid.). The research goals drive the appropriate units of analysis. If the goal would be to investigate the relationship between supernatural punishment and large-scale cooperation, this could be conceptualized at both the individual-level and the culture-level. However, encoding data above the individual-level inherently reduces variation. This is particularly problematic when investigating large-scale societies, which exhibit more cultural variation than smaller-scale societies. Forcing a binary value on whether a society has moralizing gods present/absent is dubious when there is substantial variation in individual beliefs. Furthermore, Slingerland et al. (2020) recommend that uncertainty about variables ought to be incorporated in the structure of the database. Uncertainty about variables may arise due to sparseness of sources or due to disagreement between experts. A way to capture this uncertainty is to allow for degrees of uncertainty by incorporating value ranges (e.g., population estimates) or coder confidence ratings (ibid.). As such, the database should not force single values upon variables.

4.1.4. Galton's problem

Functional relationships in cross-cultural research may be confounded due to cultures often not being statistically independent from one another due to shared historical lineages and cultural contact. This challenge is also known as 'Galton's problem' (e.g., Ember and Ember, 2009). As Watts et al. (2015) argue, many studies that investigate the relationship between moralizing gods and social complexity rely on correlational tests that do not sufficiently correct for statistical non-independence. Galton's problem is a pervasive challenge, and cross-cultural studies must negotiate it.

However, there is no straightforward way to account for this problem. For example, the SCCS was explicitly designed to contend with this challenge by sampling geographically distant cultures (Murdock and White, 1969). Despite these efforts, Dow and Eff (2008) find significant levels of autocorrelation still in the SCCS, meaning that it does not sufficiently control for non-independence of cultures. To be fair, though, as all human cultures are related, some degree of autocorrelation is always to be expected. Furthermore, Brown and Eff (2010) have determined that nearly a quarter of the variation in moralizing gods can be attributed to cultural diffusion. Turchin et al. (2022) attempt to deal with this challenge by identifying ten world regions and three 'Natural Geographic Areas', leading to 30 distinct areas in which societies can be categorized. The goal was to minimize historical relationships between cultures while simultaneously maximizing variability in the sample. Some other databases, such as the electronic Human Relations Area Files (eHRAF) attempt to account for Galton's problem by offering a sub-setting option called 'Probability Sample Files', which stratifies all data in the eHRAF into 60 culture areas. Subsequently, one culture is randomly selected from each of the culture areas, leading to a diverse sample. A different approach is taken by Watts et al. (2015), who claim that Bayesian phylogenetic methods provide the most powerful way of dealing with Galton's problem. This is because phylogenetic methods are able to infer independent evolutionary events instead of relying on correlations. These methods rely on the availability of robust language phylogenies, which can then be used for tracking e.g., the evolution of supernatural concepts in language. The challenge is that such phylogenies are often not available or incomplete (Evans et al., 2021; Watts et al., 2015). In sum, there is no standardized way of negotiating Galton's problem.

4.1.5. Causal and statistical inference

How can we ascertain whether supernatural punishment engenders large-scale cooperation or the other way around? Or is there perhaps a feedback loop whereby large-scale cooperation and supernatural punishment co-evolve? Identifying causal directionality is one of the key challenges to the debate (Turchin et al., 2022). Most studies in this debate have relied on correlational designs, making it impossible to properly distinguish the direction of causality (Watts et al., 2015).

For example, the main point of Whitehouse et al. (2022) is that socially complex societies

precede Big Gods. Their article is a critique of the 'Big Gods'-hypothesis, which states that beliefs in supernatural enforcement facilitated the earliest increased in social complexity (Norenzayan, 2013). Therefore, Whitehouse et al. proceed to argue that, if this hypothesis were true, we would find that Big Gods precede socially complex societies. This is because Big Gods are a supposed cause of such societies. However, the scholars find the opposite: increases in social complexity actually precede the development of Big Gods. Therefore, they argue that Big Gods did not contribute to the development of social complexity in human history.

This causal interpretation is wrong on several accounts. First, it is a variation of the '*post hoc ergo propter* hoc'-fallacy. Just because A occurred after B, does not mean that B caused A. Applied to our case, just because social complexity did NOT follow Big Gods does not necessarily mean that Big Gods played no causal role in the emergence of social complexity. This is because such emergence may be caused by several factors apart from Big Gods.

Second, Norenzayan's Big Gods account argues that 'Big Gods were *one* critical causal factor that contributed to the rise of large groups unleashed by agriculture' (2013, 120-1, my emphasis). In other words, there are multiple causes of large groups besides Big Gods. One implication of this is that societies with Big Gods do not necessarily have to be large-scale because there are potentially other factors at play that prevent the upscaling of societies (such as ecological or agricultural factors). Similarly, this also means that large-scale societies do not necessarily have to feature Big Gods because these gods are not a necessary feature of large-scale societies.

Third, Whitehouse et al. (2022) fail to take into account the data generation process. As they employ the Seshat cross-cultural database which relies on codings of ethnographic data, their data analysis is mediated through these sources. Therefore, the presence or absence of Big Gods is conditional on whether these were recorded in the source material in the first place. As Lightner et al. (2023) and Purzycki and McKay (2023) argue, the presence/absence of moralizing gods is highly conditional on the presence of writing in a society as well as other biases in data collection, which leads to a crucial underestimation of the presence of moralizing gods in small-scale societies (that often lack writing). Therefore, the attempted falsification of the Big Gods hypothesis is causally unsound.

Another recent example of causal misinterpretations can be found in Turchin et al. (2022). For any kind of causal inference to be valid, a causal model should be designed *before* the data analysis stage (McElreath, 2020). Turchin et al. create their causal model *on the basis of* their data analysis. This is an untenable approach because causality can never be established by the data alone (ibid.). Furthermore, the authors confuse prediction with causation. They claim to employ a causal paradigm called 'Granger causality', which they refer to as evolutionary causality in their paper, and use dynamic

regression analyses to show how a change in one variable can predict another. However, labeling this as establishing causality is misleading because a prediction does not necessarily reflect an underlying causal reality. This is because prediction is merely concerned with estimating future outcomes based on observed patterns. On the other hand, causation speaks to the mechanistic relationships between causes and effects. Essentially, using this methodology, variables are explained by virtue of how their past values relate to current values and how the past values of other variables relate to it. However, this does not reflect any causal structure; it merely reflects the predictive association between variables. Furthermore, Turchin et al. (2022) use Akaike's information criterion to determine the 'best' possible model. However, this approach ignores causal inference, and solely focusses on predictive quality. Additionally, it favors confounded models (McElreath, 2020).

In summary, some evolutionary studies of religion face challenges in correctly applying causal inference, which emphasizes the need of foregrounding it.

4.2. Addressing Methodological Challenges

This thesis has been primarily concerned with crafting different causal models related to supernatural punishment and cooperation. These models are explicitly based on the assumptions spelled out by theory. However, if we would like to base future research designs on these models, we will also need to account for Galton's problem as well as the data generation process. This is because these factors can possible confound the relationship between supernatural punishment and cooperation. This section makes suggestions on how we can model Galton's problem and the nature of missing data. I also suggest some possible research directions based on the causal models in the previous section, and I provide a table of potential operationalizations to advance conceptual and operational clarification.

4.2.1. Modelling Galton's problem

To reiterate briefly, Galton's problem entails that cultures can be hard to compare due to cultural contact and shared evolutionary and cultural history (Ember and Ember, 2009). For example, extensive cultural contact (be it through trade, colonization or other factors) influences the development of religion as well as norms and institutions that form the basis of cooperation. What this means in practice is that we should be careful in making strong universalist claims regarding the nature of supernatural punishment and cooperation if the sample does not adequately negotiate cultural

contact and shared cultural evolutionary history. This is because supernatural punishment might as well be the product of evolutionary history instead of socio-ecological variables such as warfare or ecological risk. To illustrate the confounding potential of Galton's problem, I display a simplified DAG below:

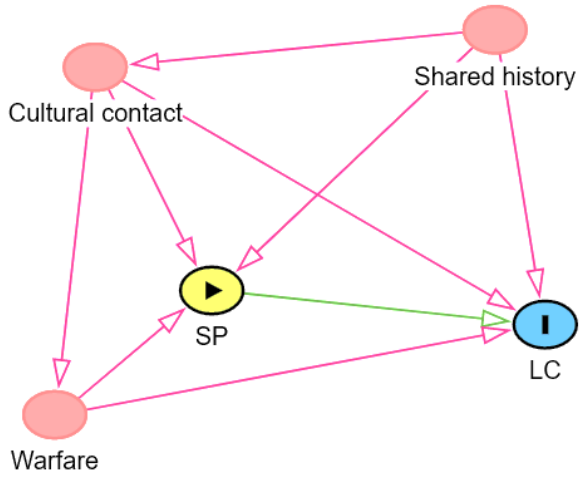

Figure 17. An example model of how the processes that induce Galton's problem can be included in the causal structure.

The DAG shows how both cultural contact and shared evolutionary history confounds the relationship between supernatural punishment and large-scale cooperation. Therefore, it is essential that these variables are held constant in large cross-cultural comparative studies if the goal is to make high quality causal inferences.

Sampling geographically independent societies has been the main strategy in contending with Galton's problem (such as the sampling strategy of SCCS, the Probability Sample Files of HRAF, and the NGA of Seshat). only Bayesian phylogenetic methods are unequivocally capable of dealing with Galton's problem (Watts et al., 2015). However, such methods are often impractical and hard to implement (Evans et al., 2021; McElreath, 2020). The most common strategies of negotiating Galton's problem are mostly practical in nature: in large-scale ethnographic collaborative projects, the team of researchers may be confined by their ethnographic expertise (Purzycki et al., 2022), meaning that the sample of cultures may be sub-optimal in regard to controlling for the effects of cultural contact and shared evolutionary history between the sampled cultures. Similarly, large-scale cross-cultural databases are reliant on pre-existing ethnographic records, which entails that they may not reflect an ideal sample because these records may not comprise a representative sample of the world's cultures. For example, the sampling strategy of the SCCS was specifically designed to contend with Galton's problem, but still significant levels of autocorrelation remain (Dow and Eff, 2008). Therefore, the field

may progress by reassessing how it deals with Galton's problem.

One possible solution is to explicitly model the effects of cultural contact and shared cultural evolutionary history (cf. Watts et al., 2022). For example, to contend with the problem of shared evolutionary histories, we may consider the migration patterns of human's earliest ancestors to account for how and where they spread. In addition, we may consider language families that potentially reflect these migration patterns, an explanatory strategy often used in archaeology and linguistics (McConvell, 2010). For example, Bouckaert et al. (2012), argued using phylogeographic approaches that the origin of the Indo-European language family spread from Anatolia with the expansion of agricultural practices 8000-9500 years ago. To implement such ideas in practice, we could stratify by or sample from different language families to hold the effect of shared evolutionary history constant. For example, large cross-cultural studies could limit themselves to one cultural group per language family.

An additional possibility is to focus on cultural groups speaking isolated languages. Such languages have no known relationships with other languages, or they are the last surviving language of a language family (Campbell, 2010), which may have been due to these cultures developing in isolation for an extended period of time (Urban, 2021). However, even cultures featuring isolated languages can feature extensive cultural contact, which can lead to acculturation and even assimilation (ibid.). For example, Korean, Japanese and Chinese are all considered isolated languages, yet their cultural histories are characterized by considerable cultural contact, which inevitably has had an important influence on their respective cultural developments.

Therefore, to fully contain Galton's problem, we also need an index of cultural contact. To do so, we would need some operationalizable measure of cultural contact. The most commonly used in cross-cultural databases is geographic proximity. The idea is that societies in closer proximity most likely have a higher rate of contact than those far away from each other. This is mostly true for small-scale societies, as larger-scale state-level societies will still find ways to engage in cultural contact due to the globalized nature of our world. Therefore, in addition to geographic proximity, many other measures may also be possible, for example the (estimated) amount of trade interactions, history of conquest (e.g., colonial influences, Christian missionaries) and migration rate. The challenge here lies in gathering all of this additional data, which may be unfeasible or even impossible in certain situations. However, as mentioned, controlling for geographic proximity for small-scale societies may already get the job done.

Thus, assuring a diverse cross-cultural sample capable of negotiating Galton's problem requires us to consider the shared evolutionary histories (by e.g., stratifying by language families) as well as the rate of cultural contact between the sampled cultures (by e.g., sampling from

geographically distant societies, and/or including measures of trade interactions, migration rate, and the extent of colonial influence and history of conquest). For practical reasons, confining comparative research projects to small-scale societies may be most fruitful as sampling from geographically distant societies may already sufficiently control for cultural contact (e.g., it is highly likely that a small-scale tribe in the Amazon will never have had contact with small-scale tribes in other continents). Although challenging, this is not an impossible endeavor.

4.2.2. Modelling Missing data

Issues related to the data-generation process can potentially bias the inferences we get to make. How is the data that we get to analyze produced, and through which factors is it mediated? Similarly, what could cause data to be missing? Whether missing data biases inference depends on the nature of the missingness. Missing values are always induced by some kind of process, and this process may be relevant to the causal model (McElreath, 2020). If, for example, the missing data are missing completely at random (meaning that missingness is unrelated to the outcome variable), then this will not lead to a biased estimate (McElreath, 2020). However, missing data on supernatural punishment and social complexity in the current ethnographic record is highly non-random. For example, small-scale societies systematically have less recorded data on moralizing gods (Lightner et al., 2023). Purzycki et al. (2022) argue that unobserved but true social complexity influences the development of writing systems (see figure 18). Subsequently, this influences observable data because scholars rely on such written sources. Socially complex societies are more likely to have writing systems as opposed to small-scale societies that often lack writing. Indeed, McElreath (2020) finds that 84% of missing values in his dataset are from non-literate societies. This leads to a systematic underestimation of the presence of moralizing gods that monitor and punish selfish behavior (Lightner et al., 2023).

Furthermore, missionaries may also underreport the presence of such gods due to biases discussed earlier, such as the 'primitive monotheism'-bias. Such reports can cause both missing data as well as measurement error. Below I display a DAG that represents the causal influences of a potential data-generation process in a large-scale ethnographic context (from Purzycki et al., 2022).

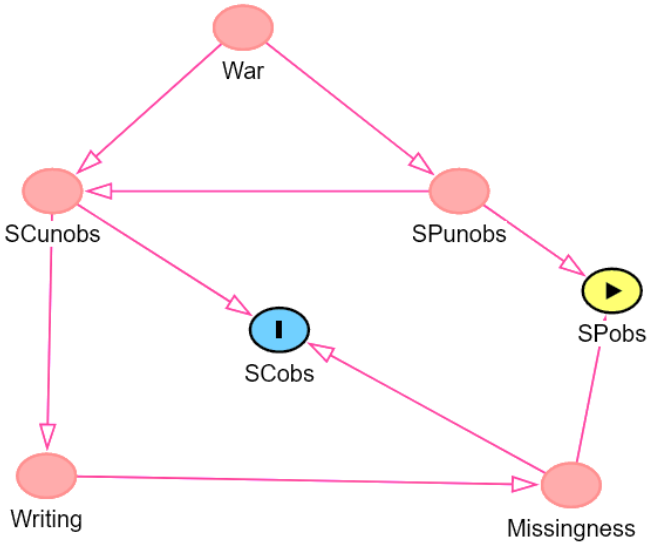

Figure 18. The causal model from Purzycki et al. (2022) shows how the causal nature of missing data could create insurmountable challenges in the statistical analysis.

The only way to resolve the DAG is to hold constant missingness and either true but unobserved social complexity or supernatural punishment. However, there is no data on unobserved social complexity or supernatural punishment. The only possible way to generate that data is through imputation, whereby missing data are replaced by a simulated value based on the available information. However, imputation requires us to have a good model and understanding by which the unobserved data is generated (McElreath, 2020). Currently, this information is just unavailable, making this particular problem statistically unresolvable. That said, whether missing data induces a problem in statistical analysis depends on the features of the dataset. Therefore, ideally, we would have to construct a cross-cultural database in which the missing data is not contingent on or associated with other important variables in the model. All this is to say that we need to carefully consider the processes that went into the creation of the data that we ultimately get to analyze.

4.2.3. Potential future research directions

A prevalent challenge in this scientific discussion (and any social scientific endeavor) is to draft a workable operational design that provides an adequate test of evolutionary hypotheses concerning religion. In an ideal scenario, this design would flow from the conceptual and theoretical framework. However, often we are limited by the availability of the data, which may lead to scholars to use suboptimal variables for the concepts they wish to represent. As I argued, this has been the case for contemporary evolutionary theories of religion, where much work has been done through pragmatic motivations by virtue of the availability of the data (e.g., the SCCS and EA being the few large cross-cultural databases available for a long time). This is a common reality in research, and the best way to

deal with this is to be honest and upfront about the limitations of the design of the study. Unfortunately, not many scholars do this. A notable exception is Johnson (2005, 418), who mentions in his SCCS study that:

> The ideal variable for this study would be a measure of the extent of belief in supernatural punishment for selfishness within each society. Unfortunately, no such variable exists in the SCCS database.

He then goes on to admit (420):

> Some caveats are in order at the outset. Even if the variable "high gods" precisely equated with the extent of belief in supernatural punishment from them (which it probably does not), it cannot be a perfect index of expected punishment for norm transgressions as a whole. This is because among the diversity of world cultures: (1) not all high gods are expected to punish all transgressions; (2) not all supernatural punishment is attributed to high gods and (3) not all punishment, of course, is supernatural: transgressors may suffer worldly punishment from real people as well.

In the last decade or so, scholars have begun to push for purpose-built databases, which resulted in large cross-cultural databases such as Seshat, the Database for Religious History, and Pulotu (Turchin et al., 2022; Slingerland and Sullivan, 2017; Watts et al., 2015). These databases were built with specific research questions and theoretical frameworks in mind. This has been a pivotal development for the field. No longer are we bound to imperfect proxies of the variables we wish to investigate, and we are instead investigating theoretically relevant variables.

Another possible way forward to is to consider the morality-as-cooperation framework of Curry (2016). Based on this framework, we could include also other measures of cooperation beside resource sharing to see weather supernatural punishment influences the whole breadth of cooperation. This is because supernatural punishment may increase resource sharing for example, but it is unclear whether this applies to other forms of cooperation, such as labor division, as well. Studies probing the relationship between large-scale cooperation and supernatural punishment have so far relied on limited conceptualizations of cooperation. However, cooperation is a multi-faceted concept. By taking Curry et al.'s framework, we can discern at least 7 different kinds of cooperation (corresponding to a form of morality). If we wish to understand the full breadth of the influence of supernatural punishment on cooperation (or the other way around), it will be useful to account for these different kinds of cooperation. Of course, it is not necessary for studies to take all of these into account due to practical reasons. However, if studies would only focus on cooperation defined in terms of resource sharing, we will be missing the broader picture.

To increase operational transparency, I have compiled a table of potentially theoretically relevant concepts in the supernatural punishment-cooperation constellation, as well as the operationalizations that have been applied so far in the literature (not exhaustive). Additionally, I have made suggestions of my own.

Concept	**Operational comparison and suggestions**
Supernatural Punishment	1) 'The extent of belief in supernatural punishment for selfishness' (Johnson, 2005). 2) Belief in moralizing (punitive and monitoring) gods that care about norm transgressions (Lang et al., 2019) 3) "A supernatural agent or process that reliably monitors and punishes selfish actions, and this concept must (i) be widely advocated within the community, (ii) involve punishment of a broad range of selfish behaviors and (iii) apply to a wide range of community members". (Watts et al., 2015) As Johnson (2015) suggests, it may be useful to distinguish between supernatural punishment in the afterlife and this life, future studies may also wish to include measures that take that into account (see e.g., Turchin et al., 2022).
Supernatural reward	Extent to which cooperative behaviors are rewarded by a supernatural agent. (Johnson, 2015)
(Large-scale) cooperation	Large-scale cooperation generally concerns social behaviors susceptible to free-riding in interactions between anonymous strangers,

	characterized by unclear guarantee of reciprocity (Johnson, 2005). Economic games can be used in individual-level studies (e.g., Lang et al., 2019; see also Pisor et al., 2020). Johnson (2005, 420-421) uses the following proxy measures, arguing that cultural groups with supernatural punishment will be: 1) Larger, since their success in achieving cooperative pursuits will have allowed them to expand, avoid fission, and compete successfully with other societies 2. More compliant with social norms and decisions 3. More able to lend money and use abstract media of exchange, since this requires high degrees of trust and guarantees (Swanson 1960) 4. More loyal to the local and wider community 5. More sharing with food 7. More likely to pay taxes, since people may be more willing to contribute to the public good 8. Less likely to experience internal conflict, if common moralizing regulations bind the society together in common cause. Curry et al.'s morality-as-cooperation framework (2016) discerns 7 kinds of cooperation, which could be considered for future research. These include: 1) resource allocation, 2) Coordination to mutual advantage,

	3) social exchange, 4) conflict resolution through displays of either dovish or (5) hawkish traits, 6) division and 7) possession. Future research may tap these kinds of cooperation by applying economic games.
Resource scarcity	Exposure to more abundant rainfall, higher primary productivity, and greater biodiversity (Botero et al., 2014; Spicer et al., 2022).
Intergroup competition	Presence of a variety of military technology, presence of cavalry (Turchin et al., 2022; Bellah, 2011). Henrich et al. (2019) employ a war exposure index, which uses all of the available information for each individual by summing the dichotomous answers to all available questions on war-violence (experienced, witnessed or perpetrated) and loss of property.
Agricultural productivity	Modes of subsistence ranging from hunter-gatherers, horticulturalists, pastoralists, agriculturalists (Peoples and Marlowe, 2012; Lang et al., 2019). Tons of carbohydrate produced (wheat, rice, maize, root vegetables, etc.) per hectare per year (Turchin et al., 2022; Baumard et al., 2013)
Ecological risk	Exposure to predictable annual cycles of precipitation and temperature as well as warmer and stable temperature (Botero et al., 2014). Temperature and precipitation reconstructions may be feasible as well, although temperature reconstructions are more likely to be inaccurate (Spicer et al., 2022).

Human punishment	The presence or absence of human punishment systems can be operationalized in terms of presence of police force, law enforcement system, or legal system, see e.g., Lang et al., 2019). Beyond presence/absence measurements, further investigations can be conducted to assess the effectiveness and societal regard for these punitive authorities.
Cultural contact	Geographic proximity, history of conquest and imperialism, migration rate are examples of potential variables that could be used to control for cultural contact.
Shared evolutionary history	In attempting to hold constant the influence of shared cultural evolutionary history, we might consider stratifying by language families. This is because language families reflect such shared history (McConvell, 2010). Additionally, historical migration patterns can provide cues about the cultural history of societies. Two societies will be more autocorrelated if they are the product of a shared migration history.

I now suggest some potential research directions based on the causal models and the table above. For each model, I develop some ideas about how an ideal test of each model would look like, holding constant the confounding influences caused by Galton's problem (i.e., cultural contact and shared evolutionary history) to the best of our ability.

As specified in the previous chapter, the SPH model only features four variables: supernatural punishment, human punishment, (large-scale) cooperation, and supernatural reward. We might take Johnson's ideal operationalization of supernatural punishment and measure the extent to which gods punish selfish behaviors. Such behaviors could be experimentally tested using economic games, such as the RAG or DG. However, other suggestions following Curry's morality-as-cooperation framework may be progress the field more as these other conceptions of cooperation have not yet been tested in the evolutionary science of religion. For example, other kinds of economics games, such as coordination games, tap a different domain of cooperation (Cooper and Weber, 2020). Furthermore,

holding constant supernatural reward is optional, but if desired, it may be measured as the extent to which gods supernaturally reward cooperative behaviors. Finally, a measure is required for human punishment as it can play a role both in causing supernatural punishment and large-scale cooperation. One way to do so is to control for the presence/absence of some kind of police force or any other secular institution tasked with punishing selfish behaviors.

The prosocial religions model is untestable if the target relationship is to estimate the influence of supernatural punishment on large-scale cooperation. There are two alternatives. One is that the reverse association may be considered, where large-scale cooperation contributes to the spread of supernatural punishment. In that case, we would need to hold constant the variables agricultural productivity, human punishment, group size, and intergroup competition. Thus, an empirical project could for example focus on subsistence modes, presence/absence of secular punishment institutions, an approximation of group size, and a war exposure index (following Henrich et al., 2019). The second alternative is to consider the co-evolutionary nature of the model, but this would require multiple time points for every variable.

The HBE model is concerned with explaining how supernatural punishment manifests itself as an evolved response to local socio-ecological challenges, such as warfare, cooperation, and ecological stress. The model that I developed provided two ways of estimating the effect of large-scale cooperation on supernatural punishment. The first way to estimate this effect is to hold constant the confounding variables agricultural productivity, ecological risk, resource scarcity, and warfare. The second way is to hold constant agricultural productivity, warfare, and group size. For example, an empirical research project seeking to understand how supernatural punishment evolves as a response to large-scale cooperation might take into account tons of carbohydrate produced per year, presence of military technology or cavalry, an estimate of group size, and an economic game reflecting cooperative behavior.

5. Discussion

The primary objective of this thesis has been to compare and contrast several theoretical frameworks regarding the causal nature of the relationship between supernatural punishment and cooperation. In doing so, I explicated a series of causal models based on these frameworks to determine if the research problems could be resolved using the logic of causal inference. In doing so, I identified the variables necessary for controlling confounding factors. I conducted causal simulations to demonstrate the potential impact of confounding. Finally, I suggested possible roadmaps for future studies.

I discerned three main hypotheses that have been posited in the literature. The first is that supernatural punishment stimulates large-scale cooperation (the SPH model). The second is that supernatural punishment is an adapted response to socio-ecological pressures (the HBE model). The third is that supernatural punishment and large-scale cooperation co-evolve in a feedback loop (the prosocial religions model). This chapter first elucidates the relevance of the causal analysis and embeds it within the literature before proceeding to provide general recommendations for future research. I close with some final remarks on how this thesis contributes to social science at large.

5.1. Interpretation and Implications of the Causal Analysis

From the theoretical literature, I derived a series of possible causal models. Additionally, I ran simulations to show to what extent 1) predefined estimates are retrievable, and 2) how confounding could impact the retrieved estimate.

The aim of the first goal was to show that it is possible to achieve an unbiased result when investigating the relationship between supernatural punishment and large-scale cooperation in the cases of the SPH model and the HBE model. I also showed that it is impossible to retrieve an unbiased estimate of the effect of supernatural punishment on large-scale cooperation for the prosocial religions model. The influence of large-scale cooperation on supernatural punishment, however, is retrievable for the prosocial religions model.

The aim of the second goal was to show how not acknowledging the effects of confounding variables can lead to biased results, impacting scientific inference. This is a crucial point for future empirical research in the field, as much published literature lacks careful causal designs. Most papers do not consider enough how potential confounding factors may influence their results. Even if they do, they do not employ causal models of any kind. While an abundance of theories and hypotheses exist that attempt to explain the relation between supernatural punishment and large-scale cooperation, not enough studies take an explicit causal approach to test these hypotheses. Studies often appear to discuss each other's hypotheses as competing claims (e.g., Turchin et al., 2022; Whitehouse et al., 2022), whereas I argue that they are different sides of the same coin — the most important differences lie in whether supernatural punishment (or large-scale cooperation) is considered the explanans or the explanandum. Thus, the field may progress by carefully engaging with causal hypotheses through established causal inference methods.

Of course, this does not mean that the causal models I outlined should be regarded as definitive or complete. That said, DAGs are crude and should never be considered complete or definitive. This is because hypotheses usually do not translate to DAGs one-to-one, and hypotheses

usually imply multiple possible DAGs (McElreath, 2020). Still, the DAGs I outlined in this thesis provide a crucial first step in formalizing the hitherto verbal models on the evolution of supernatural punishment and cooperation. Prior to this study, the literature lacked this level of formalization, underscoring the relevance and contribution of this thesis.

5.2. General recommendations

Doing large-scale cross-cultural science is extremely difficult. For one, the interdisciplinary nature of this type of research contributes to this difficulty, as it includes disciplines ranging from social psychology, cognitive science of religion, evolutionary social science, quantitative ethnography, to data science and statistics. Additionally, this thesis has argued that explicit causal designs are necessary to tease apart the directions of causality of the variables involved. I argue that we can further increase progress in the field by adopting theoretical and conceptual workflows developed recently in the philosophy of psychology (Eronen and Bringman, 2021; Bringman et al., 2022).

I argue that the primary way to improve large-scale cross-cultural science starts from the bottom-up, at the level of the data (cf. Lightner et al., 2023; Purzycki and McKay, 2023; Watts et al., 2022). This is because 'good models are hard to build on the basis of bad data' (Freedman, 1985, 345). Furthermore, Bringman et al. (2022) point out that conceptual clarity, which includes proper measurement and validation, is essential for getting good data. For too long, the field has relied on insufficiently defined concepts such as 'moralizing (high) gods' and 'social/cultural/political complexity' that further obscure rather than clarify the phenomena under investigation. This is also the reason why I preferred to stick with the supernatural punishment and large-scale cooperation concepts — supernatural punishment because it directly addresses the relevant mechanism (i.e., the actual punishment has a presumed causal effect), whereas 'moralizing gods' is more broad and obscure — and large-scale cooperation is a more clearly defined concept following from evolutionary theory and game theory, whereas 'social complexity' is rarely, if ever, used in a theoretically defined sense.

Additionally, such conceptual confusion can undermine theories and hinder theoretical progress. As Wilshire et al., (2021, 336) argue: 'there is little point in continuing to develop and refine statistical techniques or classification schemes until we have a better grasp of these key concepts'. This is also the reason why we first need to achieve conceptual clarity before moving on to mathematically formalizing our theories. This is because theoretical models will not be able to accurately capture the basic conceptual structure if there is conceptual ambiguity (Bringman et al., 2022). Therefore, the causal models I sketched in this thesis most likely do not correctly reflect reality because the theories from which they are derived are informed by ambiguous concepts and poor data quality.

To proceed, Eronen and Bringmann (2021) suggest to employ 'epistemic iteration'. This entails that concepts and theories are continuously refined based on repeated cycles of revision of prior beliefs. The goal is that progress may be achieved through each epistemic cycle. To illustrate how such a cycle is borne out in science, Eronen and Bringmann offer the example of the electron, which had a different meaning when it was first conceived as opposed to how it is conceived today. Its meaning has evolved through repeated experimentation and theoretical development. For example, early physicists proposed the existence of the electron as a sub-atomic particle with negative charge. Repeated experimentation lead to the discovery of novel properties. After the quantum revolution, it was argued that electrons ought to have wave-like properties, which was then borne out in scientific experimentation. As such, the understanding and meaning of the electron concept could be refined iteratively (ibid.). Therefore, it may be fruitful to apply epistemic iteration to evolutionary studies of religion as well.

To put this into practice, I suggest the following scientific workflow for further evolutionary studies of religion, based on Bringmann et al. (2022) and McElreath (2020).

1) Research ought to go through the whole iterative cycle for every research project. This means that researchers should revisit the phenomena that are central to their explanatory goals: what do we mean with supernatural punishment and how robust is this phenomenon?
2) Conceptual ambiguities should be explicitly addressed, as well as different ways of addressing the key concepts. What different conceptions of social complexity exist in the literature? How does the social complexity concept relate to the (large-scale) cooperation concept? How does the supernatural punishment concept relate to the moralizing gods concept?
3) These conceptual definitions should then be explicitly related to the measurement methods. These methods should always be explicitly theoretically justified. If studies use the social complexity concept, how does the operationalization (e.g., levels of jurisdictional hierarchy) follow logically and/or theoretically from that definition? How have other studies operationalized this concept? Could differences in operationalization between studies be a cause for differing results? How is cooperation measured, and how do other studies measure it and why?

Thus, to advance the field, researchers should adopt an iterative cycle that continuously refines concepts and theories The suggested workflow involves addressing conceptual ambiguities, relating them to measurement methods, and establishing clear definitions. Yet, gathering high quality ethnographic data using validated measures related to religion and cooperation is already a hugely arduous task. Still, such steps are necessary to achieve meaningful scientific progress.

Another way for individual-level targeted ethnographic work to progress is to adopt a wider

range of economic games, such as coordination games, to gain a broader and deeper perspective of the influence of supernatural punishment on cooperation (e.g., Cooper and Weber, 2020). One such game we may implement are weak-link games, which are a kind of coordination game where the overall outcome depends on the performance of the weakest participant of the group (Cartwright et al., 2013). The individuals in the group share a common goal or task. The crucial characteristic is that the participants are required to coordinate and cooperate toward this common goal or task —if one participant fails to contribute in a sufficient way, the entire group's success will be undermined. A typical observation in the game theory literature is that coordination failure occurs when groups exceed three individuals (Camerer, 2003). Yamagashi and Sato (1986) argue that this failure occurs due to levels of trust falling. Supernatural punishment could intervene in this context in several ways. It may increase levels of trust among the group's members by harnessing a shared religious identity (cf. Lang et al., 2019). Additionally, supernatural punishment may raise the cost of not coordinating such that coordination is stimulated. Thus, the economic game theory literature provides a vast and standardized methodological toolset aligned with theories of cooperation that can be adapted to local cultural contexts (Pisor et al., 2020).

5.3. Final remarks

Evolutionary theories of religion have seen rapid developments over the last few decades. Earlier studies in the last century relied mostly on older sociological theories of the relationship between religion and cooperation. In recent decades, such social scientific ideas have been increasingly integrated with evolutionary theories on cultural evolution and human behavioral ecology. Furthermore, the statistical arsenal has increased in power, and the advent of causal inference allows us to test causal hypotheses at a deeper level beyond the descriptive and correlational.

In spite of that, most empirical efforts probing the relationship between supernatural punishment and large-scale cooperation are still severely limited by the quality of the available data. Additionally, the field has recently gone through statistical and causal controversies. This thesis has shown that the different theoretical models differ in the explanandum/explanans, yet these models are not mutually exclusive to one another. Therefore, scholars should refrain from emphatically arguing that their findings 'disprove' certain hypotheses (e.g., Turchin et al., 2022). Furthermore, future studies ought to exercise caution using cross-cultural databases that are primarily constructed using old missionary reports and other forms of biased data characterized by low validity. A way forward is to develop cross-cultural databases from the ground up with theoretically-driven goals in mind using data collected by modern experts. A second way forward is through targeted ethnographic work where a broader set of economic games are applied to tap the full domain of cooperation. This

is important because the quality of inference is pivotally premised on quality data.

This lesson can be extended to other social scientific ventures. The key way to make progress in social science is for scholars to be extremely clear about how they go about operationalizing their constructs in a valid manner. The hypothesis should be appropriately modelled, and the central concepts should be clarified. Finally, all of the previous mentioned suggestions are fully contingent upon the quality of the data as well as the biases responsible for the data generation. This might prove a difficult task, but we have all the tools at our disposal. For the social sciences to progress and mature, this is how high we need to set the bar.

6. Bibliography

Alexander, Richard D. *The Biology of Moral Systems*. Foundations of Human Behavior. Hawthorne, N.Y: A. de Gruyter, 1987.

Axelrod, Robert, and William D. Hamilton. 'The Evolution of Cooperation'. *Science* 211, no. 4489 (1981): 1390–96.

Baimel, A., C. L. Apicella, Q. Atkinson, A. Bolyanatz, E. Cohen, C. Handley, J. Henrich, et al. 'Material Insecurity Predicts Greater Commitment to Moralistic and Less Commitment to Local Deities: A Cross-Cultural Investigation'. *Religion, Brain & Behavior* 12, no. 1–2 (2022): 4–17.

Barton, C Michael. 'Complexity, Social Complexity, and Modeling'. *Journal of Archaeological Method and Theory* in press (1 June 2013). https://doi.org/10.1007/s10816-013-9187-2.

Baumard, Nicolas, and Pascal Boyer. 'Explaining Moral Religions'. *Trends in Cognitive Sciences* 17, no. 6 (1 June 2013): 272–80. https://doi.org/10.1016/j.tics.2013.04.003.

Baumard, Nicolas, Alexandre Hyafil, Ian Morris, and Pascal Boyer. 'Increased Affluence Explains the Emergence of Ascetic Wisdoms and Moralizing Religions'. *Current Biology* 25, no. 1 (5 January 2015): 10–15. https://doi.org/10.1016/j.cub.2014.10.063.

Beheim, Bret, Quentin D. Atkinson, Joseph Bulbulia, Will Gervais, Russell D. Gray, Joseph Henrich, Martin Lang, et al. 'Treatment of Missing Data Determined Conclusions Regarding Moralizing Gods'. *Nature* 595, no. 7866 (July 2021): E29–34. https://doi.org/10.1038/s41586-021-03655-4.

Bellah, Robert N. *Religion in Human Evolution: From the Paleolithic to the Axial Age*. Cambridge, Mass: Belknap Press of Harvard University Press, 2011.

Bendixen, Theiss, Aaron D. Lightner, Coren Apicella, Quentin Atkinson, Alexander Bolyanatz, Emma Cohen, Carla Handley, et al. 'Gods Are Watching and so What? Moralistic Supernatural Punishment across 15 Cultures'. *Evolutionary Human Sciences*, 12 May 2023, 1–30. https://doi.org/10.1017/ehs.2023.15.

Bendixen, Theiss, Aaron Lightner, and Benjamin Grant Purzycki. 'The Cultural Evolution of Religion and Cooperation'. PsyArXiv, 11 October 2021. https://doi.org/10.31234/osf.io/fhscv.

Bocquet-Appel, Jean-Pierre. 'When the World's Population Took Off: The Springboard of the Neolithic Demographic Transition'. *Science* 333, no. 6042 (29 July 2011): 560–61. https://doi.org/10.1126/science.1208880.

Boehm, Christopher. 'A Biocultural Evolutionary Exploration of Supernatural Sanctioning'. *Evolution of Religion: Studies, Theories, and Critiques*, 2008, 143–52.

Botero, Carlos, Beth Gardner, Kathryn Kirby, Joseph Bulbulia, Michael Gavin, and Russell Gray. 'The Ecology of Religious Beliefs'. *Proceedings of the National Academy of Sciences* 111 (10 November 2014). https://doi.org/10.1073/pnas.1408701111.

Bouckaert, Remco, Philippe Lemey, Michael Dunn, Simon J. Greenhill, Alexander V. Alekseyenko, Alexei J. Drummond, Russell D. Gray, Marc A. Suchard, and Quentin D. Atkinson. 'Mapping the Origins and Expansion of the Indo-European Language Family'. *Science* 337, no. 6097 (24 August 2012): 957–60. https://doi.org/10.1126/science.1219669.

Bourrat, Pierrick, and Hugo Viciana. 'Supernatural Beliefs and the Evolution of Cooperation'. In *The Oxford Handbook of Evolutionary Psychology and Religion*, edited by James R. Liddle and Todd K. Shackelford, 0. Oxford University Press, 2021. https://doi.org/10.1093/oxfordhb/9780199397747.013.23.

Bringmann, Laura F., Timon Elmer, and Markus I. Eronen. 'Back to Basics: The Importance of Conceptual Clarification in Psychological Science'. *Current Directions in Psychological Science* 31, no. 4 (1 August 2022): 340–46. https://doi.org/10.1177/09637214221096485.

Brown, Christian, and E. Anthon Eff. 'The State and the Supernatural: Support for Prosocial Behavior'. *Structure and Dynamics* 4, no. 1 (7 January 2010). https://doi.org/10.5070/SD941003308.

Brown, Donald E. *Human Universals*. Philadelphia: Temple University Press, 1991.

Bulbulia, Joseph, Uffe Schjoedt, John H. Shaver, Richard Sosis, and Wesley J. Wildman. 'Causal Inference in Regression: Advice to Authors'. *Religion, Brain & Behavior* 11, no. 4 (2 October 2021): 353–60. https://doi.org/10.1080/2153599X.2021.2001259.

Camerer, Colin F. *Behavioral Game Theory: Experiments in Strategic Interaction*. Behavioral Game Theory: Experiments in Strategic Interaction. New York, NY, US: Russell Sage Foundation, 2003.

Campbell, Lyle. 'Language Isolates and Their History, or, What's Weird, Anyway?' *Annual Meeting of the Berkeley Linguistics Society* 36, no. 1 (2010): 16.

Chudek, Maciej, and Joseph Henrich. 'Culture-Gene Coevolution, Norm-Psychology and the Emergence of Human Prosociality'. *Trends in Cognitive Sciences* 15, no. 5 (May 2011): 218–26. https://doi.org/10.1016/j.tics.2011.03.003.

Cooper, David J., and Roberto A. Weber. 'Recent Advances in Experimental Coordination Games'. In *Handbook of Experimental Game Theory*, 149–83. Edward Elgar Publishing, 2020. https://www.elgaronline.com/display/edcoll/9781785363320/9781785363320.00014.xml.

Curry, Oliver. 'Morality as Cooperation: A Problem-Centred Approach', 27–51, 2016. https://doi.org/10.1007/978-3-319-19671-8_2.

Curry, Oliver Scott, Daniel Austin Mullins, and Harvey Whitehouse. 'Is It Good to Cooperate?: Testing the Theory of Morality-as-Cooperation in 60 Societies'. *Current Anthropology* 60, no. 1 (2 February 2019): 47–69. https://doi.org/10.1086/701478.

Davis, William David. *Societal Complexity and the Nature of Primitive Man's Conception of the Supernatural*. University of North Carolina at Chapel Hill, 1971.

Deffner, Dominik, Julia M. Rohrer, and Richard McElreath. 'A Causal Framework for Cross-Cultural Generalizability'. *Advances in Methods and Practices in Psychological Science* 5, no. 3 (1 July 2022): 25152459221106370. https://doi.org/10.1177/25152459221106366.

Dovidio, John F. 'Helping Behavior and Altruism: An Empirical and Conceptual Overview'. In *Advances in Experimental Social Psychology*, edited by Leonard Berkowitz, 17:361–427. Academic Press, 1984. https://doi.org/10.1016/S0065-2601(08)60123-9.

Dow, Malcolm M., and E. Anthon Eff. 'Global, Regional, and Local Network Autocorrelation in the Standard Cross-Cultural Sample'. *Cross-Cultural Research* 42, no. 2 (May 2008): 148–71. https://doi.org/10.1177/1069397107311186.

Edel, Abraham. 'Anthropology and Ethics in Common Focus'. *The Journal of the Royal Anthropological Institute of Great Britain and Ireland* 92, no. 1 (1962): 55–72. https://doi.org/10.2307/2844321.

Ellison, George. 'The Strengths and Weaknesses of Directed Acyclic Graphs (Dags) as Cognitive, Analytical and Educational Tools for Medical Statistics'. Preprints, 8 October 2022. https://doi.org/10.20944/preprints202210.0084.v1.

Ember, Carol R., and Melvin Ember. *Cross-Cultural Research Methods*. 2nd ed. Lanham: Altamira Press, 2009.

Ember, Carol R., Ian Skoggard, Erik J. Ringen, and Megan Farrer. 'Our Better Nature: Does Resource Stress Predict beyond-Household Sharing?' *Evolution and Human Behavior* 39 (2018): 380–91. https://doi.org/10.1016/j.evolhumbehav.2018.03.001.

Eronen, Markus I., and Laura F. Bringmann. 'The Theory Crisis in Psychology: How to Move Forward'. *Perspectives on Psychological Science* 16, no. 4 (1 July 2021): 779–88. https://doi.org/10.1177/1745691620970586.

Evans, Cara L., Simon J. Greenhill, Joseph Watts, Johann-Mattis List, Carlos A. Botero, Russell D. Gray, and Kathryn R. Kirby. 'The Uses and Abuses of Tree Thinking in Cultural Evolution'. *Philosophical Transactions of the Royal Society B: Biological Sciences* 376, no. 1828 (5 July 2021): 20200056. https://doi.org/10.1098/rstb.2020.0056.

Freedman, David A. 'Statistics and the Scientific Method'. In *Cohort Analysis in Social Research: Beyond the Identification Problem*, edited by William M. Mason and Stephen E. Fienberg, 343–66. New York, NY: Springer, 1985. https://doi.org/10.1007/978-1-4613-8536-3_11.

Ge, Erhao, Yuan Chen, Jiajia Wu, and Ruth Mace. 'Large-Scale Cooperation Driven by Reputation, Not Fear of Divine Punishment'. *Royal Society Open Science* 6 (21 August 2019): 190991. https://doi.org/10.1098/rsos.190991.

Geertz, Armin W. 'Do Big Gods Cause Anything?' *Religion* 44, no. 4 (2 October 2014): 609–13. https://doi.org/10.1080/0048721X.2014.937052.

Gintis, Herbert. 'The Evolution of Private Property'. *Journal of Economic Behavior & Organization* 64, no. 1 (September 2007): 1–16. https://doi.org/10.1016/j.jebo.2006.02.002.

Hamilton, W. D. 'The Evolution of Altruistic Behavior'. *The American Naturalist* 97, no. 896 (September 1963): 354–56. https://doi.org/10.1086/497114.

Henrich, Joseph, Michal Bauer, Alessandra Cassar, Julie Chytilová, and Benjamin Grant Purzycki. 'War Increases Religiosity'. *Nature Human Behaviour* 3, no. 2 (February 2019): 129–35. https://doi.org/10.1038/s41562-018-0512-3.

Henrich, Joseph, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, Herbert Gintis, Richard McElreath, et al. '"Economic Man" in Cross-Cultural Perspective: Behavioral Experiments in 15 Small-Scale Societies'. *Behavioral and Brain Sciences* 28, no. 6 (December 2005): 795–815. https://doi.org/10.1017/S0140525X05000142.

Henrich, Joseph, Jean Ensminger, Richard McElreath, Abigail Barr, Clark Barrett, Alexander Bolyanatz, Juan Camilo Cardenas, et al. 'Markets, Religion, Community Size, and the Evolution of Fairness and Punishment'. *Science* 327, no. 5972 (19 March 2010): 1480–84. https://doi.org/10.1126/science.1182238.

Henrich, Joseph, Richard McElreath, Abigail Barr, Jean Ensminger, Clark Barrett, Alexander Bolyanatz, Juan Cardenas, et al. 'Costly Punishment Across Human Societies'. *Science (New York, N.Y.)* 312 (1 July 2006): 1767–70. https://doi.org/10.1126/science.1127333.

Henrich, Joseph, and Michael Muthukrishna. 'The Origins and Psychology of Human Cooperation'. *Annual Review of Psychology* 72, no. 1 (2021): 207–40. https://doi.org/10.1146/annurev-psych-081920-042106.

Hruschka, Daniel, Charles Efferson, Ting Jiang, Ashlan Falletta-Cowden, Sveinn Sigurdsson, Rita McNamara, Madeline Sands, Shirajum Munira, Edward Slingerland, and Joseph Henrich. 'Impartial Institutions, Pathogen Stress and the Expanding Social Network'. *Human Nature (Hawthorne, N.Y.)* 25, no. 4 (December 2014): 567–79. https://doi.org/10.1007/s12110-014-9217-0.

Johnson, Dominic. *God Is Watching You: How the Fear of God Makes Us Human*. Oxford, New York: Oxford University Press, 2015.

Johnson, Dominic, and Jesse Bering. 'Hand of God, Mind of Man: Punishment and Cognition in the Evolution of Cooperation'. *Evolutionary Psychology* 4, no. 1 (1 January 2006): 147470490600400130. https://doi.org/10.1177/147470490600400119.

Johnson, Dominic D. P. 'God's Punishment and Public Goods: A Test of the Supernatural Punishment Hypothesis in 186 World Cultures'. *Human Nature* 16 (2005): 410–46. https://doi.org/10.1007/s12110-005-1017-0.

———. 'The Wrath of the Academics: Criticisms, Applications, and Extensions of the Supernatural Punishment Hypothesis'. *Religion, Brain & Behavior* 8, no. 3 (3 July 2018): 320–50. https://doi.org/10.1080/2153599X.2017.1302986.

Johnson, Dominic D.P. 'Big Gods, Small Wonder: Supernatural Punishment Strikes Back'. *Religion, Brain & Behavior* 5, no. 4 (2 October 2015): 290–98. https://doi.org/10.1080/2153599X.2014.928356.

Johnson, Dominic, and Oliver Krüger. 'The Good of Wrath: Supernatural Punishment and the Evolution of Cooperation'. *Political Theology* 5, no. 2 (10 February 2004): 159–76. https://doi.org/10.1558/poth.2004.5.2.159.

Kappeler, Peter M. 'A Framework for Studying Social Complexity'. *Behavioral Ecology and Sociobiology* 73, no. 1 (19 January 2019): 13. https://doi.org/10.1007/s00265-018-2601-8.

Kreidler, Marc. 'Reinvigorating the Comparative, Cooperative Ethnographic Sciences of Religion | Free Inquiry', 3 April 2018. https://secularhumanism.org/2018/04/cont-reinvigorating-the-comparative-cooperative-ethnographic-sciences-of-re/.

Laidlaw, James. 'For an Anthropology of Ethics and Freedom'. *The Journal of the Royal Anthropological Institute* 8, no. 2 (2002): 311–32.

Lang, Martin, Benjamin G. Purzycki, Coren L. Apicella, Quentin D. Atkinson, Alexander Bolyanatz, Emma Cohen, Carla Handley, et al. 'Moralizing Gods, Impartiality and Religious Parochialism across 15 Societies'. *Proceedings of the Royal Society B: Biological Sciences* 286, no. 1898 (13 March 2019): 20190202. https://doi.org/10.1098/rspb.2019.0202.

Larson, Jennifer, Harvey Whitehouse, Pieter Francois, Daniel Hoyer, and Peter Turchin. *Moralizing Supernatural Punishment and Reward: A Response to Critics*, 2022. https://doi.org/10.31235/osf.io/hs5mk.

Laurin, Kristin, Azim Shariff, Joseph Henrich, and Aaron Kay. 'Outsourcing Punishment to God: Beliefs in Divine Control Reduce Earthly Punishment'. *Proceedings. Biological Sciences / The Royal Society* 279 (23 May 2012): 3272–81. https://doi.org/10.1098/rspb.2012.0615.

Lewis, David Kellogg. *Convention: A Philosophical Study*. Cambridge, MA, USA: Wiley-Blackwell, 1969.

Lightner, Aaron D., Theiss Bendixen, and Benjamin Grant Purzycki. 'Moralistic Supernatural Punishment Is Probably Not Associated with Social Complexity'. *Evolution and Human Behavior*, 23 November 2022. https://doi.org/10.1016/j.evolhumbehav.2022.10.006.

Mcconvell, Patrick. 'The Archaeo-Linguistics Of Migration'. In *Migration History in World History*, 153–86. Brill, 2010. https://brill.com/display/book/edcoll/9789004186453/Bej.9789004180314.i-287_007.xml.

McElreath, Richard. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. 2nd ed. CRC Texts in Statistical Science. Boca Raton: Taylor and Francis, CRC Press, 2020.

McKay, Ryan, and Harvey Whitehouse. 'Religion and Morality'. *Psychological Bulletin* 141, no. 2 (March 2015): 447–73. https://doi.org/10.1037/a0038455.

Montgomery, Jacob M., Brendan Nyhan, and Michelle Torres. 'How Conditioning on Posttreatment Variables Can Ruin Your Experiment and What to Do about It'. *American Journal of Political Science* 62, no. 3 (2018): 760–75. https://doi.org/10.1111/ajps.12357.

Murdock, George P., and Douglas R. White. 'Standard Cross-Cultural Sample'. *Ethnology* 8, no. 4 (October 1969): 329. https://doi.org/10.2307/3772907.

———. 'Standard Cross-Cultural Sample: On-Line Edition', 19 July 2006. https://escholarship.org/uc/item/62c5c02n.

Murdock, George Peter. *Ethnographic Atlas*. Pittsburgh: University of Pittsburgh Press, 1967. http://digital.library.pitt.edu/cgi-bin/t/text/text-idx?c=pittpress;view=toc;idno=31735057895306.

Murray, Michael J. 'The Evolution of Religion: Adaptationist Accounts'. In *Science and Religion in Dialogue*, 437–57. John Wiley & Sons, Ltd, 2010. https://doi.org/10.1002/9781444317350.ch34.

Naroll, Raoul. 'Galton's Problem: The Logic of Cross-Cultural Analysis'. *Social Research* 32, no. 4 (1965): 428–51.

Norenzayan, Ara. *Big Gods: How Religion Transformed Cooperation and Conflict*. Princeton, NJ: Princeton Univ. Press, 2013.

Norenzayan, Ara, and Azim F. Shariff. 'The Origin and Evolution of Religious Prosociality'. *Science* 322, no. 5898 (3 October 2008): 58–62. https://doi.org/10.1126/science.1158757.

Norenzayan, Ara, Azim F. Shariff, Will M. Gervais, Aiyana K. Willard, Rita A. McNamara, Edward Slingerland, and Joseph Henrich. 'Parochial Prosocial Religions: Historical and Contemporary Evidence for a Cultural Evolutionary Process'. *The Behavioral and Brain Sciences* 39 (January 2016): e29. https://doi.org/10.1017/S0140525X15000655.

———. ‘The Cultural Evolution of Prosocial Religions’. *Behavioral and Brain Sciences* 39 (ed 2016). https://doi.org/10.1017/S0140525X14001356.

Nunnally, Jum C. *Psychometric Theory*. McGraw-Hill, 1978.

Ozono, Hiroki, Nobuhito Jin, Motoki Watabe, and Kazumi Shimizu. ‘Solving the Second-Order Free Rider Problem in a Public Goods Game: An Experiment Using a Leader Support System’. *Scientific Reports* 6, no. 1 (9 December 2016): 38349. https://doi.org/10.1038/srep38349.

Pasek, Michael, Crystal Shackleford, Allon Vishkin, Julia Smith, and Jeremy Ginges. ‘Thinking About God Encourages Prosociality Toward Religious Outgroups: A Cross-Cultural Investigation’. *Psychological Science*, 18 April 2023. https://doi.org/10.1177/09567976231158576.

Pearl, Judea. ‘Causal Diagrams for Empirical Research’. *Biometrika* 82, no. 4 (1995): 669–88. https://doi.org/10.2307/2337329.

Pearl, Judea, Madelyn Glymour, and Nicholas P. Jewell. *Causal Inference in Statistics: A Primer*. Chichester, West Sussex: Wiley, 2016.

Peoples, Hervey C., Pavel Duda, and Frank W. Marlowe. ‘Hunter-Gatherers and the Origins of Religion’. *Human Nature* 27, no. 3 (1 September 2016): 261–82. https://doi.org/10.1007/s12110-016-9260-0.

Peoples, Hervey C., and Frank W. Marlowe. ‘Subsistence and the Evolution of Religion’. *Human Nature* 23, no. 3 (1 September 2012): 253–69. https://doi.org/10.1007/s12110-012-9148-6.

Peregrine, Peter. ‘The Birth of the Gods Revisited: A Partial Replication of Guy Swanson’s (1960) Cross-Cultural Study of Religion’. *Cross-Cultural Research* 30, no. 1 (1 February 1996): 84–112. https://doi.org/10.1177/106939719603000104.

Pfattheicher, Stefan, Yngwie Asbjørn Nielsen, and Isabel Thielmann. ‘Prosocial Behavior and Altruism: A Review of Concepts and Definitions’. *Current Opinion in Psychology* 44 (1 April 2022): 124–29. https://doi.org/10.1016/j.copsyc.2021.08.021.

Pisor, Anne C., Matthew M. Gervais, Benjamin G. Purzycki, and Cody T. Ross. ‘Preferences and Constraints: The Value of Economic Games for Studying Human Behaviour’. *Royal Society Open Science* 7, no. 6 (June 2020): 192090. https://doi.org/10.1098/rsos.192090.

Powers, Simon T., Carel P. van Schaik, and Laurent Lehmann. ‘Cooperation in Large-Scale Human Societies—What, If Anything, Makes It Unique, and How Did It Evolve?’ *Evolutionary Anthropology: Issues, News, and Reviews* 30, no. 4 (2021): 280–93. https://doi.org/10.1002/evan.21909.

Purzycki, Benjamin G., Daniel N. Finkel, John Shaver, Nathan Wales, Adam B. Cohen, and Richard Sosis. ‘What Does God Know? Supernatural Agents’ Access to Socially Strategic and Non-Strategic Information’. *Cognitive Science* 36, no. 5 (2012): 846–69. https://doi.org/10.1111/j.1551-6709.2012.01242.x.

Purzycki, Benjamin Grant, Coren Apicella, Quentin D. Atkinson, Emma Cohen, Rita Anne McNamara, Aiyana K. Willard, Dimitris Xygalatas, Ara Norenzayan, and Joseph Henrich. ‘Moralistic Gods, Supernatural Punishment and the Expansion of Human Sociality’. *Nature* 530, no. 7590 (February 2016): 327–30. https://doi.org/10.1038/nature16980.

Purzycki, Benjamin Grant, Theiss Bendixen, and Aaron D. Lightner. ‘Coding, Causality, and Statistical Craft: The Emergence and Evolutionary Drivers of Moralistic Supernatural Punishment Remain

Unresolved'. *Religion, Brain & Behavior* 0, no. 0 (29 June 2022): 1–8. https://doi.org/10.1080/2153599X.2022.2065349.

Purzycki, Benjamin Grant, Theiss Bendixen, Aaron D. Lightner, and Richard Sosis. 'Gods, Games, and the Socioecological Landscape'. *Current Research in Ecological and Social Psychology* 3 (1 January 2022): 100057. https://doi.org/10.1016/j.cresp.2022.100057.

Purzycki, Benjamin Grant, Joseph Henrich, Coren Apicella, Quentin D. Atkinson, Adam Baimel, Emma Cohen, Rita Anne McNamara, Aiyana K. Willard, Dimitris Xygalatas, and Ara Norenzayan. 'The Evolution of Religion and Morality: A Synthesis of Ethnographic and Experimental Evidence from Eight Societies'. *Religion, Brain & Behavior* 8, no. 2 (3 April 2018): 101–32. https://doi.org/10.1080/2153599X.2016.1267027.

Purzycki, Benjamin Grant, Martin Lang, Joseph Henrich, and Ara Norenzayan. 'Guiding the Evolution of the Evolutionary Sciences of Religion: A Discussion'. *Religion, Brain & Behavior* 12, no. 1–2 (3 April 2022): 226–32. https://doi.org/10.1080/2153599X.2021.2021552.

———. 'The Evolution of Religion and Morality Project: Reflections and Looking Ahead'. *Religion, Brain & Behavior* 12, no. 1–2 (3 April 2022): 190–211. https://doi.org/10.1080/2153599X.2021.2021546.

Purzycki, Benjamin Grant, and Ryan McKay. 'Moralistic Gods and Social Complexity: A Brief History of the Problem'. Preprint. PsyArXiv, 14 January 2022. https://doi.org/10.31234/osf.io/vqjk2.

Purzycki, Benjamin Grant, Cody T. Ross, Coren Apicella, Quentin D. Atkinson, Emma Cohen, Rita Anne McNamara, Aiyana K. Willard, Dimitris Xygalatas, Ara Norenzayan, and Joseph Henrich. 'Material Security, Life History, and Moralistic Religions: A Cross-Cultural Examination'. Edited by Michiel van Elk. *PLOS ONE* 13, no. 3 (7 March 2018): e0193856. https://doi.org/10.1371/journal.pone.0193856.

Roes, Frans L., and Michel Raymond. 'Belief in Moralizing Gods'. *Evolution and Human Behavior* 24 (2003): 126–35. https://doi.org/10.1016/S1090-5138(02)00134-4.

Rohrer, Julia M. 'Thinking Clearly About Correlations and Causation: Graphical Causal Models for Observational Data'. *Advances in Methods and Practices in Psychological Science* 1, no. 1 (1 March 2018): 27–42. https://doi.org/10.1177/2515245917745629.

SANDERSON, STEPHEN, and WESLEY ROBERTS. 'The Evolutionary Forms of the Religious Life: A Cross-Cultural, Quantitative Analysis'. *American Anthropologist* 110 (17 December 2008): 454–66. https://doi.org/10.1111/j.1548-1433.2008.00078.x.

Sargent, R G. 'Verification and Validation of Simulation Models'. *Journal of Simulation* 7, no. 1 (1 February 2013): 12–24. https://doi.org/10.1057/jos.2012.20.

Shariff, Azim, and Ara Norenzayan. 'God Is Watching You Priming God Concepts Increases Prosocial Behavior in an Anonymous Economic Game'. *Psychological Science* 18 (1 October 2007): 803–9. https://doi.org/10.1111/j.1467-9280.2007.01983.x.

———. 'The Birth of High Gods How the Cultural Evolution of Supernatural Policing Influenced the Emergence of Complex, Cooperative Human Societies, Paving the Way for Civilization'. *Evolution, Culture, and the Human Mind*, 1 January 2011. https://doi.org/10.4324/9780203848746.

Sibley, Chris G., and Joseph Bulbulia. 'Faith after an Earthquake: A Longitudinal Study of Religion and Perceived Health before and after the 2011 Christchurch New Zealand Earthquake'. *PloS One* 7, no. 12 (2012): e49648. https://doi.org/10.1371/journal.pone.0049648.

Simpson, John H. 'High Gods and the Means of Subsistence'. *Sociological Analysis* 45, no. 3 (1984): 213–22. https://doi.org/10.2307/3711478.

Singh, Manvir, Ted Kaptchuck, and Joseph Henrich. 'Small Gods, Rituals, and Cooperation: The Mentawai Water Spirit Sikameinan'. SocArXiv, 9 September 2019. https://doi.org/10.31235/osf.io/npkdy.

Skoggard, Ian, Carol Ember, Emily Pitek, Joshua Jackson, and Christina Carolus. 'Resource Stress Predicts Changes in Religious Belief and Increases in Sharing Behavior'. *Human Nature* 31 (1 September 2020). https://doi.org/10.1007/s12110-020-09371-8.

Skyrms, Brian. *Evolution of the Social Contract*. Second edition. Cambridge ; New York: Cambridge University Press, 2014.

Slingerland, Edward, Quentin Atkinson, Carol Ember, Oliver Sheehan, Michael Muthukrishna, Joseph Bulbulia, and Russell Gray. 'Coding Culture: Challenges and Recommendations for Comparative Cultural Databases'. *Evolutionary Human Sciences* 2 (1 June 2020): 1–20. https://doi.org/10.1017/ehs.2020.30.

Slingerland, Edward, and Brenton Sullivan. 'Durkheim with Data: The Database of Religious History'. *Journal of the American Academy of Religion* 85, no. 2 (1 June 2017): 312–47. https://doi.org/10.1093/jaarel/lfw012.

Slingerland, Edward, M. Willis Monroe, Brenton Sullivan, Robyn Faith Walsh, Daniel Veidlinger, William Noseworthy, Conn Herriott, et al. 'Historians Respond to Whitehouse et al. (2019), "Complex Societies Precede Moralizing Gods Throughout World History"'. *Journal of Cognitive Historiography* 5, no. 1–2 (6 November 2020): 124–41. https://doi.org/10.1558/jch.39393.

Smith, Daniel. 'Cultural Group Selection and Human Cooperation: A Conceptual and Empirical Review'. *Evolutionary Human Sciences* 2 (ed 2020). https://doi.org/10.1017/ehs.2020.2.

Smith, J. Maynard, and G. R. Price. 'The Logic of Animal Conflict'. *Nature* 246, no. 5427 (November 1973): 15–18. https://doi.org/10.1038/246015a0.

Snarey, John. 'The Natural Environment's Impact upon Religious Ethics: A Cross-Cultural Study'. *Journal for the Scientific Study of Religion* 35 (1 June 1996): 85. https://doi.org/10.2307/1387077.

Spicer, Rachel, M. Willis Monroe, Matthew Hamm, Andrew Danielson, Gino Canlas, Ian Randall, and Edward Slingerland. 'Religion and Ecology: A Pilot Study Employing the Database of Religious History'. *Current Research in Ecological and Social Psychology* 3 (1 January 2022): 100073. https://doi.org/10.1016/j.cresp.2022.100073.

Stewart, Peter. 'Complexity Theories, Social Theory, and the Question of Social Complexity'. *Philosophy of the Social Sciences* 31, no. 3 (1 September 2001): 323–60. https://doi.org/10.1177/004839310103100303.

Sullivan, Brenton, Michael Muthukrishna, Frederick S. Tappenden, and Edward Slingerland. 'Exploring the Challenges and Potentialities of the Database of Religious History for Cognitive Historiography'. *Journal of Cognitive Historiography*, 29 March 2018, 12–31.

Swanson, Guy Edwin. *The Birth of the Gods: The Origin of Primitive Beliefs*. Nachdr. Ann Arbor Paperbacks. Ann Arbor, Mich: Univ. of Michigan Press, 1992.

Textor, Johannes, Benito van der Zander, Mark S. Gilthorpe, Maciej Liśkiewicz, and George T.H. Ellison. 'Robust Causal Inference Using Directed Acyclic Graphs: The R Package "Dagitty"'. *International Journal of Epidemiology*, 15 January 2017, dyw341. https://doi.org/10.1093/ije/dyw341.

Tinbergen, N. 'On Aims and Methods of Ethology'. *Zeitschrift Für Tierpsychologie* 20, no. 4 (1963): 410–33. https://doi.org/10.1111/j.1439-0310.1963.tb01161.x.

Trivers, Robert L. 'The Evolution of Reciprocal Altruism'. *Quarterly Review of Biology* 46, no. 1 (1971): 35–57.

Turchin, Peter, Harvey Whitehouse, Sergey Gavrilets, Daniel Hoyer, Pieter François, James S. Bennett, Kevin Feeney, et al. 'Disentangling the Evolutionary Drivers of Social Complexity in Human History: A Comprehensive Test of Hypotheses'. SocArXiv, 11 October 2021. https://doi.org/10.31235/osf.io/tekb6.

Turchin, Peter, Harvey Whitehouse, Jennifer Larson, Enrico Cioni, Jenny Reddish, Daniel Hoyer, Patrick E. Savage, et al. 'Big Gods and Big Science: Further Reflections on Theory, Data, and Analysis'. *Religion, Brain & Behavior* 0, no. 0 (29 June 2022): 1–14. https://doi.org/10.1080/2153599X.2022.2065354.

———. 'Explaining the Rise of Moralizing Religions: A Test of Competing Hypotheses Using the Seshat Databank'. *Religion, Brain & Behavior* 0, no. 0 (30 June 2022): 1–28. https://doi.org/10.1080/2153599X.2022.2065345.

Tylor, Sir Edward Burnett. *Primitive Culture: Researches Into the Development of Mythology, Philosophy, Religion, Language, Art, and Custom*. Henry Holt and Company, 1889.

Underhill, Ralph. 'Economic and Political Antecedents of Monotheism: A Cross-Cultural Study'. *American Journal of Sociology* 80, no. 4 (January 1975): 841–61. https://doi.org/10.1086/225893.

Urban, Matthias. 'The Geography and Development of Language Isolates'. *Royal Society Open Science* 8, no. 4 (n.d.): 202232. https://doi.org/10.1098/rsos.202232.

Watts, Joseph, Joseph Bulbulia, Russell D. Gray, and Quentin D. Atkinson. 'Clarity and Causality Needed in Claims about Big Gods'. *Behavioral and Brain Sciences* 39 (ed 2016): e27. https://doi.org/10.1017/S0140525X15000576.

Watts, Joseph, Simon J. Greenhill, Quentin D. Atkinson, Thomas E. Currie, Joseph Bulbulia, and Russell D. Gray. 'Broad Supernatural Punishment but Not Moralizing High Gods Precede the Evolution of Political Complexity in Austronesia'. *Proceedings of the Royal Society B: Biological Sciences* 282, no. 1804 (7 April 2015): 20142556. https://doi.org/10.1098/rspb.2014.2556.

Watts, Joseph, Joshua Jackson, Chris Arnison, Elise Hamerslag, John Shaver, and Benjamin Purzycki. 'Building Quantitative Cross-Cultural Databases From Ethnographic Records: Promise, Problems and Principles'. *Cross-Cultural Research*, 13 December 2021, 106939712110657. https://doi.org/10.1177/10693971211065720.

Watts, Joseph, Oliver Sheehan, Simon J. Greenhill, Stephanie Gomes-Ng, Quentin D. Atkinson, Joseph Bulbulia, and Russell D. Gray. 'Pulotu: Database of Austronesian Supernatural Beliefs and Practices'. *PLOS ONE* 10, no. 9 (23 September 2015): e0136783. https://doi.org/10.1371/journal.pone.0136783.

West, Stuart A., Claire El Mouden, and Andy Gardner. 'Sixteen Common Misconceptions about the Evolution of Cooperation in Humans'. *Evolution and Human Behavior* 32, no. 4 (1 July 2011): 231–62. https://doi.org/10.1016/j.evolhumbehav.2010.08.001.

Whitehouse, Harvey, Pieter François, Patrick E. Savage, Thomas E. Currie, Kevin C. Feeney, Enrico Cioni, Rosalind Purcell, et al. 'Complex Societies Precede Moralizing Gods throughout World History'. *Nature* 568, no. 7751 (April 2019): 226–29. https://doi.org/10.1038/s41586-019-1043-4.

———. 'Retraction Note: Complex Societies Precede Moralizing Gods throughout World History'. *Nature* 595, no. 7866 (July 2021): 320. https://doi.org/10.1038/s41586-021-03656-3.

Whitehouse, Harvey, Pieter François, Patrick E. Savage, Daniel Hoyer, Kevin C. Feeney, Enrico Cioni, Rosalind Purcell, et al. 'Big Gods Did Not Drive the Rise of Big Societies throughout World History'. OSF Preprints, 1 April 2021. https://doi.org/10.31219/osf.io/mbnvg.

———. 'Testing the Big Gods Hypothesis with Global Historical Data: A Review and "Retake"'. *Religion, Brain & Behavior* 0, no. 0 (29 June 2022): 1–43. https://doi.org/10.1080/2153599X.2022.2074085.

Whitehouse, Harvey, Peter Turchin, Pieter François, Patrick E. Savage, Thomas E. Currie, Kevin C. Feeney, Enrico Cioni, et al. 'A New Era in the Study of Global History Is Born but It Needs to Be Nurtured'. *Journal of Cognitive Historiography* 5, no. 1–2 (2018): 142–58. https://doi.org/10.1558/jch.39422.

Yamagishi, Toshio, and Kaori Sato. 'Motivational Bases of the Public Goods Problem'. *Journal of Personality and Social Psychology* 50, no. 1 (1986): 67–73. https://doi.org/10.1037/0022-3514.50.1.67.