An Economic Theory of Salvation Religions*

Weijia Li† Shaoda Wang‡

September 1, 2026

Abstract

This paper constructs a unified model that micro-founds the three types from Max Weber's typology of world religions, "mystic religions," "activist religions," and "religions of world adjustment." We then employ the model to study the interaction between world religions and secular institutions. Among others, the model shows that salvation religions might have played a role in the great "reversal of fortune" between the East and the West. Because China and India were the economic powerhouses in the Middle Ages, they might have failed to sustain religious activism that facilitated modern economic growth and constitutional government in Western Europe. The model further interprets competition over socioeconomic status as a partial substitution for salvation religions, explaining why status competition is fierce where the influence of salvation religions have declined.

*We thank Jean-Paul Carvalho and Xiaoming Zhang for insightful and valuable discussion of our paper. We have also benefited from thoughtful comments from Robert Akerlof, Samuel Bazzi, Sascha Becker, Roland Bénabou, Jeanet Sinding Bentzen, Alberto Bisin, Arthur Campbell, Ernesto Dal Bó, Sebastiano Della Lena, Quoc-Anh Do, Paola Giuliano, Brad Hackinen, Yu Hao, Leander Heldring, Ruixue Jia, Erik Kimbrough, Rachel Kranton, Kaveh Majlesi, Benjamin Marx, Nathan Nunn, Jean-Philippe Platteau, Giacomo Ponzetto, Pablo Querubín, Gérard Roland, Jared Rubin, Raúl Sánchez de la Sierra, Ferenc Szűcs, Ludger Wößmann, Yang Xie, Melanie Xue, and Noam Yuchtman. We are grateful for feedback from participants of talks in KCL, Liaoning, LSE, Monash, Peking, and the ASREC 2024 Conference at Chapman University, the 2025 NBER Conference on Economics of Culture and Institutions, the 2025 ERINN Conference at Warwick Venice Center, and the 2026 Workshop on Historical Political Economy in Zhejiang University. All errors are our own.

†Monash University. Email: weijia.li@monash.edu.

‡University of Chicago and NBER. Email: shaoda@uchicago.edu.

1 Introduction

Why did modern capitalism emerge in the West, not in China or India? This "Great Divergence" (Pomeranz (2000); Greif (2006)) has been attributed to different cultural orientations that have deep religious roots (Weber (1978); Tillich (1963); McCleary and Barro (2006); Alesina and Giuliano (2015); Mokyr (2016); Becker et al. (2024); Bergeron et al. (2026)). In China, Confucianism affirms and "adjusts" to the secular world, therefore unable to revolutionize the traditional economy. The secular world is devalued by Indian religions and Judeo-Christianity. But while Indian religions ultimately encourage a minimization of worldly actions to attain a "mystical escape" from the corrupt world, many denominations of Judeo-Christianity command their followers to maximize their worldly actions as an "active" instrument of God, potentially capable of revolutionizing the secular world. This typology of world religions from Max Weber is summarized by Figure 1.

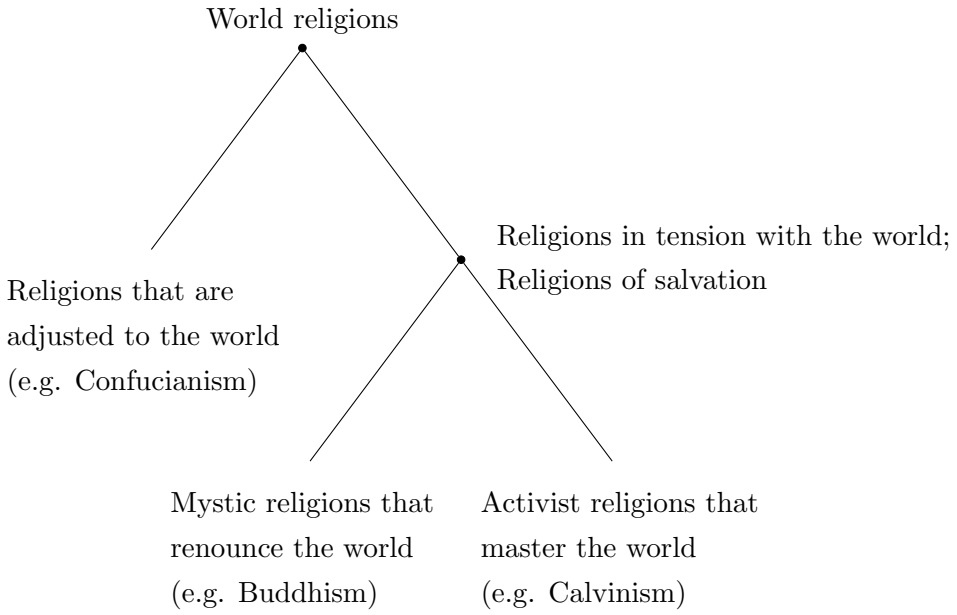


Figure 1: Weber's typology of world religions (adapted from Swedberg and Agevall (2016))

In this paper, we construct a model to jointly micro-found the three types in Max Weber's typology: "religions of world adjustment," "mystic religions," and "activist religions." Our model starts with the idea of supernatural reward and punishment that is central to many religions, as empirically documented by a large literature on Big Gods and Karma (Norenzayan (2013); Purzycki et al. (2016); White et al. (2019); White et al. (2021)). Under this idea, a follower will not suffer if he acts as demanded by his religion; but if a follower does not act as demanded, the follower will suffer from the just punishment inflicted by the divine order. Theologically correct or not, this idea of supernatural reward and punishment

is pervasive and fundamental among religions "as lived by adherents every day" (Norenzayan (2013)), the more proper domain of social scientific analysis.[1]

In this setup, one can see that salvation religions must impose commandments on human action that are extremely onerous. Under a commandment too difficult to comply, a follower never fully obeys the commandment, failing to attain the action that could have protected the follower from suffering. Such a follower can always make sense of suffering: he suffers because he acts unjustly, therefore punished by the just divine order. Here emerges a remarkable pattern. Every time the follower suffers, the traumatic experience *strengthens* his faith in his religion. These religions are not only stable against suffering. Precisely through suffering, these religions vitalize the belief that just action can attain a "salvation" from suffering.

By contrast, if a religion imposes commandments that are too moderate, the follower will eventually act as demanded by the religion. When the follower suffers again, he cannot make sense of his suffering: he has done exactly what his religion demands, yet he still suffers. The experience of suffering eventually falsifies the promise of "world-adjusting religions," which affirm action according to one's own nature.

The analysis explains why there are two opposite types of salvation religions, mysticism that demands a minimal worldly action and activism that demands a maximal worldly action (Figure 1). An individual of no religious concerns balances between the secular benefit and cost from worldly action, thus preferring an intermediate level of action. Because of this secular preference for intermediate action, a salvation religion can only sanctify either minimal or maximal action. For followers of mystic religions, full escape from worldly action would endanger their material or social existence. These followers move towards but never completely attain full escape, allowing them to attribute suffering to their lingering entanglement with the corrupt world. Analogously, followers of activist religions intensify their worldly action but never fully obey the commandment of overzealous action.

Central to our analysis, we then build on the baseline model to examine the interactions between salvation religions and secular institutions. First, the model helps explain why Judeo-Christianity tended to be far more activist than salvation religions of China and India, uncovering an economic foundation for the "religious divergence" between the East and the West (Weber (1978); Tillich (1963); Nunn (2025)). In our model, activist religions are unlikely to emerge where the material reward for worldly action is already high. The high material reward forces salvation religions to command the opposite direction of mystic escape, now an especially superhuman commandment because individuals must renounce

[1]Furthermore, our analysis focuses on the historical interactions between religions and secular institutions, when the notion of supernatural reward and punishment was probably even more prevalent than in modern data from psychologists (Belloc et al. (2016); Bentzen and Force (2023)).

the lucrative material reward from productive action. In early Middle Ages, the economy supported more lucrative surplus for merchants, officials, and cultural elites in China and India (Elvin (1973); Landes (2006); Greif et al. (2025)), potentially driving the divergence between Eastern mysticism and Western activism.[2]

This economic origin of religious divergence may shed light on the great "reversal of economic fortune" between the East and the West (Acemoglu et al. (2002); Greif et al. (2025)). Precisely because China and India had long been the economic centers of the Middle Ages, they may have failed to sustain religious activism that later revolutionized traditional economy in Western Europe. In our model, this reversal is especially likely to emerge under a strong religious motivation in search of "salvation," which may override and reverse the direction of material incentives. We therefore uncover salvation religions as a novel mediator for the great economic reversal (Figure 2).

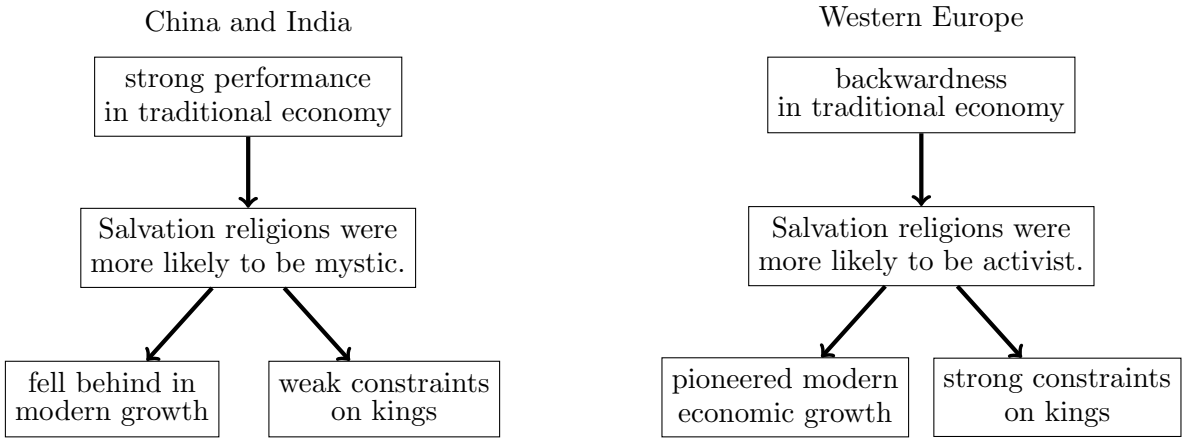


Figure 2: The potential reversal of fortune through salvation religions

Second, why did the West manage to impose strong constraints on executive power, but not historical China or India? In general, how is it ever possible to constrain executive power, which can always inflict massive sufferings on any challenger? To think about these questions, we integrate our model of salvation religions with a political game of domination, persecution, and resistance. Under intense persecution by a king, a figure of an activist religion simply concludes that he suffered because he did not do enough as an instrument of God to discipline the king. Such a belief propels an even stronger resistance against

[2]For these groups, the more lucrative surplus was obtained from the larger and denser population in China and India, where the agricultural condition were more favored and technologies were more advanced than Western Europe (Greif et al. (2025)). The living standard for peasants remained similar everywhere under the Malthusian pressure. But as Weber (2004b) has forcefully argued, peasants were rarely a reliable "carrier" of ethical religiosity. Instead, peasants were strongly attached to magic because farming relied more on natural cycles than human efforts.

the king's domination, as pioneered and exemplified by the great Hebrew prophets in their incessant struggles against political power (Finer (1997); Hayes (2012)). This remarkable transformation of suffering into determined action is rooted in the unique meaning of suffering as interpreted by activist religions.

In sharp contrast to this unwavering resistance is the eventual escapism of Chinese religio-cultural elites. To show this, we extend the model to formalize the central leitmotif of Chinese culture, the "complementarity" between Confucianism and Taoism. Since Confucianism is ultimately world-adjusting, it cannot provide enough inner support for Chinese elites who have been persecuted by the emperor. These elites will turn to the spiritual comfort of the mystic Taoism, which promises perfect equanimity out of detachment from worldly actions. Each subsequent persecution from the emperor propels these elites to escape further from their duty to discipline the emperor's power.

Finally, how could some world-affirming ideologies still be so dominant, such as Confucianism and the modern secularism (pp.14, Troeltsch (2017))? World affirmation can protect its plausibility through the alternative route of competition into elite status, promising that "salvation" is possible, but only for the elites. These status competitions can enchant "commoners" because they cannot directly test whether world affirmation *as an elite* can free one from all troubles. Our analysis may help explain the rise of status competition where the influence of salvation religions have declined, such as the Confucian civil service examination in Imperial China and many status competitions in the modern world (Abdulkadiroğlu et al. (2014); Bai and Jia (2016); Chen et al. (2020); Turchin (2023)).

The paper is related to a few strands of literature. Our model attempts to examine different religious paths to make sense of suffering, which has been identified as a central motivation of religiosity by a large literature (Norenzayan and Hansen (2006); Belloc et al. (2016); Binzel and Carvalho (2017); Hood Jr et al. (2018); Bentzen (2019); Bentzen (2021)). Our model jointly micro-founds the three Weberian types of world religions, uncovering many novel interactions between world religions and secular institutions. Among others, our model offers fresh insights on how salvation religions may have contributed to the "reversal of economic fortune" (Acemoglu et al. (2002); Greif and Tabellini (2017); Rubin (2017); Giuliano and Nunn (2021); Acemoglu and Robinson (2023); Fernández-Villaverde et al. (2023); Greif et al. (2025)).

Our analysis follows the belief approach to religion (McCleary and Barro (2006); Bénabou and Tirole (2011); Levy and Razin (2012); Augenblick et al. (2016); Nunn and Sánchez de la Sierra (2017); Bergeron et al. (2026); Espín-Sánchez et al. (2026)), complementing the club approach (Iannaccone (1992); Berman (2000); Barros and Garoupa (2002); McBride (2008); Aimone et al. (2013); Carvalho and Sacks (2021)) and the cultural transmission approach

(Bisin et al. (2004); Carvalho (2013); Bisin et al. (2024)). Adding to the many seminal insights on church strictness from existing literature, our model shows that strict commandments can protect the belief about the "redemptive" potential of such commandments. Such strictness on ethical commandments is two-sided, driving our results on the possible religion-mediated reversal of economic fortune and political divergence. Our paper also contributes to the comparative analysis of religious worldviews as endogenous equilibria. This approach is pioneered by Bénabou and Tirole (2006) on a belief in a just world, which compares agnosticism with high religiosity/"Protestant Ethic" through a different mechanism.

By focusing on religious coping with suffering, we show that religion may revitalize political resistance by endowing suffering in the hands of a tyrant with a special meaning. This offers one potential solution to the perennial puzzle of how the "soft" power of culture can possibly constrain the "hard" power of violence, contributing to the literature that examines the foundations of constitutionalism and constraints on political power (Fearon (2011); Chaney (2013); Belloc et al. (2016); Akerlof (2017); Bazzi et al. (2020); Becker and Pfaff (2023); Bazzi et al. (2026); Bisin et al. (2024)).

We also contribute to the burgeoning literature on non-Abrahmic religions (Butinda et al. (2023); Le Rossignol et al. (2025); Jia and Kung (2025)). To our knowledge, we construct the first model that formalizes the central leitmotif of Chinese culture, the complementarity between Confucianism and Taoism (Jia and Kung (2025)). This provides a religion-based explanation for why the Chinese regime has long been autocratic (Stasavage (2020); Francois et al. (2023); Egorov and Sonin (2024); Jia et al. (2024)). Our model also uncovers the decline of salvation religions as a novel explanation for the rise of status competition in various contexts (Abdulkadiroğlu et al. (2014); Chen et al. (2020)).

The paper is organized as follows. Section 2 presents the baseline model. Section 3 shows that salvation religions can reverse "economic fortune." Section 4 explores a religio-cultural foundation of executive constraints. Section 5 establishes the linkage between world affirmation and status competition. Section 6 concludes.

2 A model of salvation religions

2.1 The setup

Our objective in Section 2 is to provide a parsimonious model that can still micro-found all three Weberian types of world religions. This baseline model has an infinite horizon, with

the time index $t = 1, 2, ...$ At each period t, an individual chooses an action

$$a_t \geq 0.$$

For each period t, the individual may suffer from a disaster at a cost $K > 0$. Is it possible to be liberated from suffering? There are two competing worldviews.

2.1.1 The religious worldview versus the naturalist worldview

Under the religious worldview, action is "meaningful" in terms of explaining and avoiding suffering. At each period t, the individual suffers with probability λ_H, where

$$\lambda_H = \begin{cases} \bar{\lambda}_H \in (0,1) & \text{if and only if } a_t \neq a^h. \\ 0 & \text{if and only if } a_t = a^h. \end{cases} \tag{1}$$

Under the religious worldview, an individual does not suffer if he chooses the just action or the religious commandment at $a^h \geq 0$; but if he chooses an unjust action at $a_t \neq a^h$, the individual suffers from the punishment by the divine order with probability $\bar{\lambda}_H \in (0,1)$.[3] As discussed before, this view is sometimes contested in theological debates, but it has remained pervasive and central among followers of world religions, as empirically established by the large literature on Big Gods and Karma (e.g., Norenzayan (2013); White et al. (2021); also see Bentzen (2019) and Michalopoulos and Xue (2021)).[4]

The religious worldview competes with a "naturalist" or mechanical worldview. Under the naturalist worldview, at each period t, the individual suffers with a probability λ_M that is independent of the individual's action:

$$\lambda_M > 0 \text{ for all } a_t \geq 0. \tag{2}$$

Under this worldview, one's own action is meaningless in terms of explaining or avoiding suffering, so just and unjust people suffer equally. In addition, $\bar{\lambda}_H > \lambda_M$, so that a just divine order inflicts punishment on an unjust individual with a probability ($\bar{\lambda}_H$) higher than

[3]It is straightforward to incorporate a set of sanctified actions: under our setup, there is generically a unique optimum among a closed set of sanctified actions.

[4]Even for sophisticated followers who don't necessarily believe that disasters will cease to arrive, they still seek *internal equanimity* when they are stricken with external disasters. For these followers, Max Weber has emphasized that this "internal immunity against suffering" is a more proper concept of salvation as studied by social science rather than theology (pp.218, Weber (2004a); pp.66, Weber (2004b)). For instance, action cannot bring afterlife salvation under the Protestant doctrine of predestination. But the psychological confirmation of "certainty of grace" is still achieved primarily through action, central to the famous thesis of Protestant Ethic (Weber (1992); Troeltsch (1931)).

a morally indifferent world (λ_M).

At the start of $t = 1$, the individual believes that the religious worldview is true with a prior probability $\mu_0 \in (0, 1)$. How does the individual update his belief on the two worldviews? We assume an asymmetric process in belief updating.

1. If the individual suffers, he uses Bayes' rule to update his belief on μ_t whenever possible:

$$\mu_t = \frac{\mu_{t-1}\lambda_H}{\mu_{t-1}\lambda_H + (1 - \mu_{t-1})\lambda_M}.$$

2. If the individual does not suffer, he maintains his belief as the previous period:

$$\mu_t = \mu_{t-1}.$$

The asymmetric updating reflects a much stronger demand for "making sense" of sufferings than good fortunes. Under the asymmetric updating, the individual thoroughly uncovers all the implications of a traumatic experience for his worldviews, but he does not reflect on such existential questions when he has just experienced good fortune. The asymmetric updating resonates with the "negativity bias," a fundamental and robust finding in psychology (Baumeister et al. (2001); Rozin and Royzman (2001); Vaish et al. (2008)). Specifically, human beings display a marked "propensity to attend to, learn from, and use negative information far more than positive information" (Vaish et al. (2008)).[5]

Even though the asymmetric updating might be closer to the actual updating process, we also show that a model with a fully Bayesian individual preserves all key insights. Details are in Appendix B.

2.1.2 The payoff function

The individual chooses action $a_t \geq 0$ to maximize:

$$\underbrace{[\psi F(a_t) - C(a_t)]}_{\text{secular payoff}} + \underbrace{H(a_t - a^h; \mu_{t-1})}_{\text{religious guilt}} \underbrace{- \Big[\mu_{t-1}\mathbf{1}\{a_t \neq a^h\}\bar{\lambda}_H + (1 - \mu_{t-1})\lambda_M\Big]K}_{\text{expected cost from suffering}} \tag{3}$$

The individual's payoff has three components.

1. The first component is the secular payoff. The term $\psi F(a_t)$ is the material benefit, where the parameter $\psi > 0$ determines the material return to action. The term $C(a_t)$

[5]Psychologists usually explain the strong bias by the intense evolutionary pressure to avoid negative shocks that could bring disastrous consequences (Vaish et al. (2008)).

is the material cost of action. Define a key variable:

$$a^m(\psi) \in \arg\max_{x \geq 0}[\psi F(x) - C(x)]. \tag{4}$$

The variable $a^m(\psi)$ is the **secular optimal action** that only maximizes the secular payoff. For $a^m(\psi)$ to be unique and interior under any ψ, assume $F' > 0$, $F'' < 0$, $C' > 0$, $C'' > 0$, and $\lim_{a\to\infty} C'(a) = \infty$; in addition, we set $F(0) = \lim_{a\downarrow 0} F(a) = -\infty$, presumably because full inaction would destruct one's material and social existence.

2. The term $H(a_t - a^h; \mu_{t-1})$ is the cost of internalized guilt from disobeying the religious commandment at a^h. A parametric example of H is:

$$H(a_t - a^h; \mu_{t-1}) = -\rho(\mu_{t-1})(a_t - a^h)^2, \text{ with } \rho(0) = 0 \text{ and } \rho'(\cdot) > 0.$$

In general, assume that $H(x;\mu) \leq 0$ and $H(0;\mu) = H(x;0) = 0$. Under $\mu \neq 0$, $\partial H(x;\mu)/\partial x > 0$ for $x < 0$ and $\partial H(x;\mu)/\partial x < 0$ for $x > 0$; under $x \neq 0$, $\partial H(x;\mu)/\partial \mu < 0$. Intuitively, the individual feels more guilty under a larger distance between his action and the commandment ($|x| = |a_t - a^h|$) or under a stronger belief in religion ($\mu = \mu_{t-1}$). We also assume regular second and cross derivatives. [6]

Guilt is placed at the very center of the religious experience of salvation religions (pp.35-66, Nietzsche (1998); pp.242-243, Weber (2004a); pp.48-50, Tillich (2008); Della Lena et al. (2023)). As we will see, our model offers a novel explanation for why the feeling of guilt is so pervasive among followers of great salvation religions.

3. The third component is

$$-[\mu_{t-1} \cdot \mathbf{1}\{a_t \neq a^h\} \cdot \bar{\lambda}_H + (1 - \mu_{t-1}) \cdot \lambda_M]K, \tag{5}$$

which is the expected cost of suffering.

   (a) First, the individual believes that the religious worldview is true with a probability of μ_{t-1}. In this case, if the individual acts justly at a^h, he does not suffer; by choosing an unjust action, he suffers with a probability of $\bar{\lambda}_H$.

   (b) Second, the individual believes that the naturalist worldview is true with a probability of $1 - \mu_{t-1}$. In this case, the individual suffers with a probability λ_M, whether he acts justly or not.

[6] The second derivative $\partial^2 H(x;\mu)/\partial x^2 < 0$ for $\mu \neq 0$; the cross derivative $\partial^2 H(x;\mu)/\partial x \partial \mu > 0$ for $x < 0$ and $\partial^2 H(x;\mu)/\partial x \partial \mu < 0$ for $x > 0$. These conditions are quite general, which are true for $H(x;\mu) = -\rho(\mu)|x|^\xi$ for $x \in \mathbb{R}$ and $\mu \in [0,1]$ under $\rho(0) = 0$, $\rho'(\cdot) > 0$, and $\xi \geq 2$.

In sum, avoidance of disasters is not the only motive to act morally. As captured by the guilt function, one is also motivated by internalized moral standard. Our setup thus reflects both consequentialist and on-consequentialist motives behind ethical action.

We can also add an extra term $M(\mu_{t-1}) \geq 0$, the direct utility from believing in a "meaningful" world. It is also sensible that $M'(\mu_{t-1}) > 0$. We don't need to add this direct utility from meaningfulness, however, for all our results. To break indifference, we also assume that an individual will disobey if he is indifferent between obeying the religious commandment a^h and the disobeying action that gives the highest payoff.

Suffering in our model When our paper refers to suffering, it refers to pain that is brought by external disasters, or the term 5. We are **not** talking about, for example, the self-inflicted pain by choosing an action a_t away from the secular optimal action $a^m(\psi)$. Many salvation religions actually put a high value on such a self-inflicted pain, which results from denying material impulses (Troeltsch (1931)).

The timeline To summarize, the timeline in each period $t = 1, 2, ...$ is as follows.

1. From the last period the individual inherits μ_{t-1}, his belief in the religious worldview.
2. The individual chooses his action a_t.
3. The individual may suffer from a disaster.
4. The individual updates his belief that the religious worldview is correct, obtaining μ_t.

2.2 Analysis of the model

We first define two basic types of religions, "stable" religions versus "unstable" religions.

Definition 1.

Denote $\mu(i)$ as the belief after the individual has experienced $i \in \mathbb{N}$ sufferings.

1. *A religion is stable if and only if* $\lim_{i\to\infty} \mu(i) = 1$.
2. *A religion is unstable if and only if* $\lim_{i\to\infty} \mu(i) = 0$.

A stable religion is maximally resilient against sufferings. Even when the individual has suffered for an arbitrarily large number of times, the individual will not abandon the religion. Instead, the individual will establish a full and unwavering belief in its "salvation" promise.

From here on only stable religions are called salvation religions, since only their salvation promises are fully convincing in the long run. By contrast, an unstable religion will eventually be abandoned by the individual after a sufficiently large number of sufferings.

The following proposition characterizes stable religions.

Proposition 1. *1. A religion is stable if and only if its commandment $a^h \geq 0$ is sufficiently far away from the secular optimal action $a^m(\psi)$.*

Formally, there exist a unique cutoff $\underline{a}(\psi, \bar{\lambda}_H K) \in \big(0, a^m(\psi)\big)$ and a unique cutoff $\bar{a}(\psi, \bar{\lambda}_H K) \in \big(a^m(\psi), \infty\big)$, such that a religion a^h is stable if and only if

$$a^h \leq \underline{a}(\psi, \bar{\lambda}_H K) \text{ or } a^h \geq \bar{a}(\psi, \bar{\lambda}_H K). \tag{6}$$

Both cutoffs $\underline{a}(\psi, \bar{\lambda}_H K)$ and $\bar{a}(\psi, \bar{\lambda}_H K)$ are functions of ψ, the material return to action, and $\bar{\lambda}_H K$, suffering from disobedience as expected by the religion.

2. Under a stable religion, the individual's action moves ever closer to the commandment a^h, but the individual never exactly obeys the commandment a^h.

Formally, denote a_t^* as the individual's action at period t. Under a stable religion a^h, for all $t = 1, 2, \ldots$, $|a_{t+1}^* - a^h| \leq |a_t^* - a^h|$, and $a_{t+1}^* \neq a^h$.

All proofs are in Appendix A. A religion can only win full devotion if the religion promises "salvation" through an "extreme" commandment a^h. Suppose that the distance between a^h and $a^m(\psi)$ is large enough, i.e., $a^h \leq \underline{a}(\psi, \bar{\lambda}_H K)$ or $a^h \geq \bar{a}(\psi, \bar{\lambda}_H K)$. Under such a large distance, it is too costly for the individual to fully obey the commandment at a^h. The individual's optimal action a_t^* always balances between the secular bliss point $a^m(\psi)$ and the commandment a^h. This is expressed by the first order condition for a_t^*:

$$\underbrace{\psi F'(a_t^*) - C'(a_t^*)}_{=0 \text{ at } a^m(\psi)} + \underbrace{H'(a_t^* - a^h; \mu_{t-1})}_{=0 \text{ at } a^h} = 0$$

With a larger μ_{t-1}, the individual moves closer to a^h, driven by a stronger guilt of disobeying the commandment. But such an individual still never exactly performs a^h if the distance between a^h and $a^m(\psi)$ is sufficiently large.

Here comes the key mechanism of the model. Because the individual never exactly obeys the religious commandment, he can always make better "sense" of suffering through the religious worldview than the naturalist worldview. After suffering from a disaster, the individual reasons that he suffers because the divine order was inflicting a just punishment on a "rebellious" follower. This religious explanation is more convincing than the naturalist

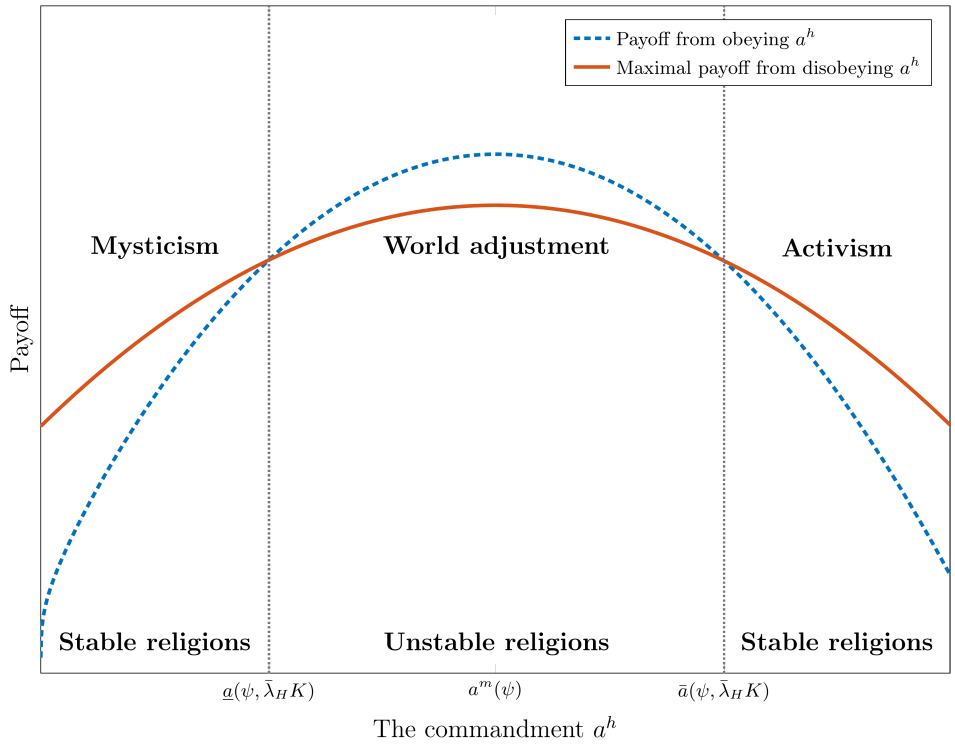


Figure 3: Religious stability and a Weberian typology of world religions

explanation because a just divine order should inflict a punishment with a higher probability on an unjust follower than a morally indifferent world. As the individual suffers, he updates his belief of worldviews, obtaining an ever stronger faith in his religion. In the long run, the individual approaches full faith. Though such an individual still fails to exactly obey the commandment a^h, his long-run action has moved as close as possible to a^h.

Remark: the centrality of guilt in religious experience and action In our model, guilt is always experienced by a follower of a salvation religion because the strict commandment generates an endogenous gap between God's commandment at a^h and human action at $a_t^* \neq a^h$. This may explain why guilt is such a crucial emotion in the religious experience of great salvation religions (pp.35-66, Nietzsche (1998); pp.242-243, Weber (2004a); pp.48-50, Tillich (2008)). Weber (2004a) indeed claims that, for followers of salvation religions, "all action in the civilized world [...] appears to be burdened with the greatest guilt" (pp. 242, Weber (2004a)).

Note that inner guilt is also indispensable for a salvation religion to guide human action. Without inner guilt ($H \equiv 0$), a religion becomes a purely intellectual exercise because even a fully believing follower still chooses $a_t^* = a^m(\psi)$ that only maximizes the secular payoff.

Proposition 1 is illustrated in Figure 3. We now characterize "unstable religions."

Proposition 2. *1. A religion is unstable if and only if it sanctifies a "moderate" action as the commandment $a^h \geq 0$:*

$$a^h \in \left(\underline{a}(\psi, \bar{\lambda}_H K), \bar{a}(\psi, \bar{\lambda}_H K)\right), \tag{7}$$

where the cutoffs $\underline{a}(\psi, \bar{\lambda}_H K)$ and $\bar{a}(\psi, \bar{\lambda}_H K)$ are the same as in Proposition 1.

2. For an unstable religion, the long-run action is the secular optimal action. Specifically, there is a finite integer $I < \infty$, so that if the individual suffers more than I disasters,

$$a_t^* = a^m(\psi).$$

When a religion sanctifies an action a^h too close to the secular optimal action $a^m(\psi)$, in the long run the individual will exactly obey the religious commandment at a^h. When he suffers again, the individual is bound to lose all his faith in the religion, fully adopting the naturalist worldview. Such an individual will choose what his human nature dictates because his action is meaningless in deliverance from suffering.

2.3 Religiosity of full Bayesians

The model is robust to full Bayesian updating, with or without disasters. Religion may even allow an obeying individual ($a_t = a^h$) to suffer at a strictly positive probability $\underline{\lambda} > 0$, though still lower than the probability in a morally indifferent world at λ_M. In this setup, Appendix B shows that the conclusions of our baseline model are robust, once we assume that the objective probability of disasters is high enough. In particular, only an unattainable commandment can induce full belief in the long run, independent of the initial belief and the sequence of disasters. Thus, our analysis is not driven by asymmetric updating, which can indeed be replaced by frequent occurrences of disasters for a Bayesian.

Our analysis also extends to a forward-looking Bayesian (Appendix C); this incorporates the motive of experimentation where the individual may "test" the religious promise to make more informed decisions in the future. Furthermore, we also consider the religiosity of fortunate individuals who experience relatively few disasters. This uncovers additional insights on the strength of strict religions, under which the religiosity of followers are more observable and "legible" than under a world-adjusting religion. We need to introduce and employ more technical tools for these analyses, so we assign them to appendices.

2.4 A micro-founded Weberian typology of world religions

We now highlight how Proposition 1 and Proposition 2 manage to micro-found Max Weber's typology of world religions through a corollary.

Corollary 1. *There are three types of religions.*

1. $a^h \leq \underline{a}(\psi, \bar{\lambda}_H K)$ *is a "mystic religion" that sanctifies a "minimal" worldly action. A follower will fully believe the religious promise, choosing a low level of action.*

2. $a^h \geq \bar{a}(\psi, \bar{\lambda}_H K)$ *is an "activist religion" that sanctifies a "maximal" worldly action. A follower will fully believe the religious promise, choosing a high level of action.*

3. $a^h \in \left(\underline{a}(\psi, \bar{\lambda}_H K), \bar{a}(\psi, \bar{\lambda}_H K)\right)$ *is a religion of "world adjustment." A follower will eventually lose faith, choosing the secular optimal action* $a^m(\psi)$.

The corollary is also illustrated in Figure 3. Our model therefore micro-founds Weber's typology of world religions in Figure 1: a world-adjusting religion that affirms human nature has too little "tension with the world," unable to offer a credible path of liberation from suffering. Such a path is only credible if a religion is in sharp "tension with the world," or a large distance between a^h and $a^m(\psi)$. This sharp tension takes two opposite directions, activist religions that master the world ($a^h > \bar{a}(\psi, \bar{\lambda}_H K) > a^m(\psi)$) and mystic religions that renounce the world ($a^h < \underline{a}(\psi, \bar{\lambda}_H K) < a^m(\psi)$). We now employ the baseline model to examine the interactions between world religions, the economy, and the political order.

3 Salvation religions and the economy

In the baseline model, the two cutoffs that separate the three types of world religions are $\underline{a}(\psi, \bar{\lambda}_H K)$ and $\bar{a}(\psi, \bar{\lambda}_H K)$, which are functions of ψ, the material return to action, and $\bar{\lambda}_H K$, suffering from disobedience as expected under the religious worldview. This section performs comparative statics with respect to ψ as the material motivation and $\bar{\lambda}_H K$ as the religious motivation. By doing so, the model might help shed light on two key questions in comparative religion and comparative development:

1. Why did Judeo-Christianity become far more activist than salvation religions of China and India, such as Buddhism, Hinduism, and Taoism?

2. Could salvation religions have played a significant role in the great reversal of economic fortune between the East and the West?

3.1 Economic foundations of salvation religions

This section shows that it might be the economic backwardness of Western Europe in the early Middle Ages that drove its salvation religions to be far more activist than Chinese or Indian ones. To demonstrate this, we investigate how the material return ψ affects the likelihood of different types of salvation religions.

Specifically, assume that the commandment a^h is a random variable. The "initial" probability density function of a^h is $l(\cdot)$ over the support $[\underline{A}^h, \bar{A}^h] \subset \mathbb{R}^+$, where $\underline{A}^h$ and $\bar{A}^h$ are the bounds on possible religious commandments. Assume that $l(x) > 0$ for all $x \in [\underline{A}^h, \bar{A}^h]$. Also, $0 \leq \underline{A}^h < \underline{a}(\psi, \bar{\lambda}_H K)$ and $\bar{a}(\psi, \bar{\lambda}_H K) < \bar{A}^h \leq \infty$. The distribution function is denoted as $L(\cdot)$, so the "final" density function for religions $a^h \notin \left(\underline{a}(\psi, \bar{\lambda}_H K), \bar{a}(\psi, \bar{\lambda}_H K)\right)$ is:

$$\frac{l(a^h)}{1 - [L(\bar{a}(\psi, \bar{\lambda}_H K)) - L(\underline{a}(\psi, \bar{\lambda}_H K))]}.$$

Thus, for an unstable religion with

$$a^h \in \Big(\underline{a}(\psi, \bar{\lambda}_H K), \bar{a}(\psi, \bar{\lambda}_H K)\Big),$$

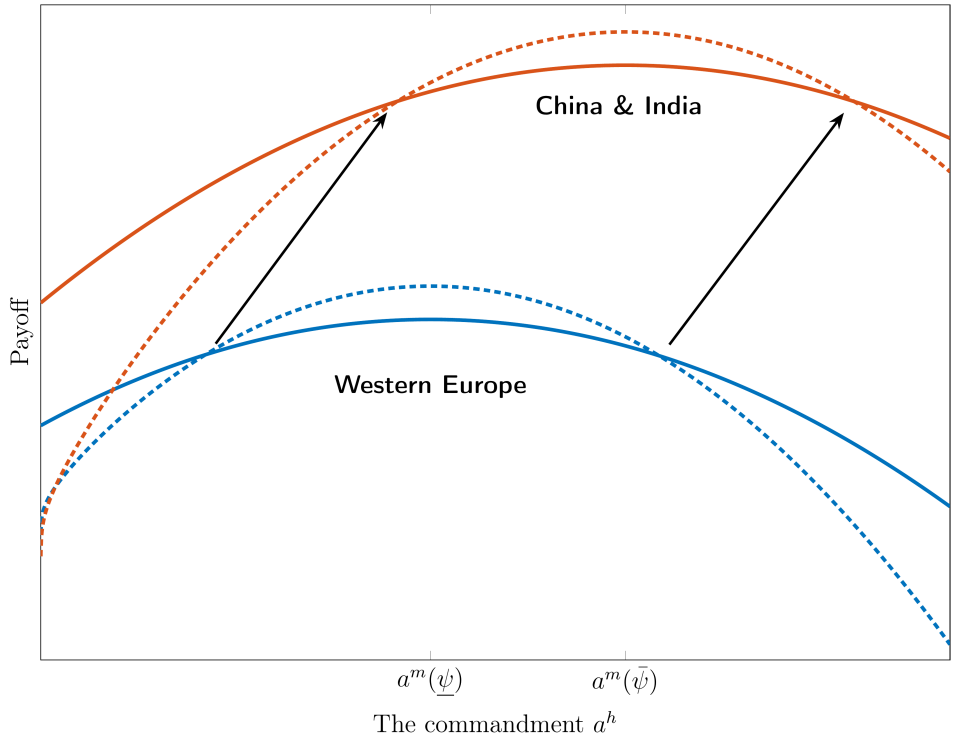
its followers will eventually be disenchanted. These followers will be matched with a stable religion. The "probability" of matching with a specific stable religion a^h is determined by its density $l(a^h)$.[7] We can now characterize an economic foundation of salvation religions.

Proposition 3. *Under a higher material return ψ, a salvation religion a^h is more likely to be mystic and less likely to be activist.*

The proposition is visualized by Figure 4. We obtain Proposition 3 by showing that a higher material return ψ drives up both cutoffs $\underline{a}(\psi, \bar{\lambda}_H K)$ and $\bar{a}(\psi, \bar{\lambda}_H K)$. Therefore, the set of mystic religions $[\underline{A}^h, \underline{a}(\psi, \bar{\lambda}_H K)]$ expands, while the set of activist religions $[\bar{a}(\psi, \bar{\lambda}_H K), \bar{A}^h]$ shrinks. By doing so, we demonstrate that a higher material return induces a higher likelihood of mystic religions and a lower likelihood of activist religions.

Proposition 3 is driven by the fundamental principle of our model, the principle that salvation religions strengthen their "salvation" promise by keeping a sufficient distance from worldly impulse. To maintain such a distance, salvation religions tend to be mystic precisely when worldly action is productive. In this case, it is a superhuman commandment to renounce the lucrative return from worldly action. Followers of mystic religions can only partially escape from lucrative action, so they can always attribute their suffering to their hesitation to detach from a corrupt world that still offers such rich benefits. This protects the

[7] This process of re-enchantment is micro-founded later in Section 4.



Note: the figure shows changes when the material return to action increases from $\underline{\psi}$ to $\bar{\psi} > \underline{\psi}$. Solid curves are maximal payoffs from disobedience and dashed curves are payoffs from obedience. Arrows show the changes to the cutoffs that separate stable religions from unstable religions.

Figure 4: Salvation religions in the East and the West

promise of liberation from suffering through mystic escape. A similar logic drives salvation religions to be activist precisely when worldly action produces few material benefits.

Strength in traditional economy and the great religious divergence Proposition 3 may help us understand the great religious divergence between the East and the West. To be sure, activism and mysticism were both present in the East and the West: there were prominent mystic movements in the long history of Judeo-Christianity, and many activist movements arose out of Buddhism and Taoism. But still, salvation religions were more likely to be mystic in China and India *in comparison with* salvation religions in Western Europe (Weber (1978)).

Our model conjectures an economic origin of the religious divergence: mysticism was more prevalent in China and India partly because they were the centers of traditional economy in the Middle Ages (Abu-Lughod (1989); Landes (2006); Greif et al. (2025)); by contrast, there were more activist denominations in Judeo-Christianity partly because Western Europe had

long been on the economic periphery.[8]

3.2 Salvation religions can reverse "economic fortune"

This section shows that salvation religions might reverse the "economic fortune" between the East and the West, a process that might be pioneered by Western locations where religious activism is particularly strong.

Consider an economic outcome Y_t that is an increasing function of action a_t:

$$Y_t = Y(a_t), \text{ with } Y'(\cdot) > 0. \tag{8}$$

Denote a^* as the optimal action of a fully believing individual ($\mu_{t-1} \to 1$) under a stable religion a^h. Action a^* is determined by the first order condition:

$$\psi F'(a^*) - C'(a^*) + H'(a^* - a^h; 1) = 0. \tag{9}$$

Equation 9 generates action $a^*(\psi, a^h)$ as a function of the material return ψ and the commandment a^h, and a^* increases with a^h. The corresponding economic outcome

$$Y(a^*(\psi, a^h)) \tag{10}$$

is therefore also a function of ψ and a^h. We will show that when the religious motivation is strong ($\bar{\lambda}_H K$ is large), it is especially likely that the economic outcome will be higher under *lower* material return.

Specifically, we compare two particular locations in the East and the West, whose material returns in traditional economy are $\underline{\psi}$ and $\bar{\psi}$, with $\underline{\psi} < \bar{\psi}$. The salvation religion is mystic in this Eastern location and activist in this Western location. Denote the religion in the Western location as $\hat{a}^h$ and the religion in the Eastern location as $\tilde{a}^h$. We can write down their density functions, which are assumed to be independent of each other:

$$\begin{cases} \text{West: } \hat{a}^h \sim \frac{l(x)}{1-L(\bar{a}(\underline{\psi},\bar{\lambda}_H K))} \text{ for } x \in [\bar{a}(\underline{\psi},\bar{\lambda}_H K), \bar{A}^h]; \\ \text{East: } \tilde{a}^h \sim \frac{l(x)}{L(\underline{a}(\bar{\psi},\bar{\lambda}_H K))} \text{ for } x \in [\underline{A}^h, \underline{a}(\bar{\psi},\bar{\lambda}_H K)]. \end{cases} \tag{11}$$

[8] As discussed before, economic surplus was higher for the non-peasant population in China and India, such as merchants, officials, and the cultural elite. The living standard for the peasants were similar everywhere, but because farming activities relied so much on natural cycles, peasants were strongly attracted to magic, not ethical religions. It was instead the non-peasant population that usually developed strong ethical religiosity (Weber (1978); Weber (2004b)).

Assume $Y(a^*(\bar{\psi}, \underline{A}^h)) < Y(a^*(\underline{\psi}, \bar{A}^h))$ so that the economic outcome Y in the Eastern location under the lowest commandment is strictly lower than the economic outcome in the Western location under the maximal commandment. We can establish the following result, formalizing the insight that a strong religious motivation can reverse economic fortune.

Proposition 4. *The probability that $Y(a^*(\bar{\psi}, \tilde{a}^h)) < Y(a^*(\underline{\psi}, \hat{a}^h))$ increases with $\bar{\lambda}_H K$.*

In other words, when individuals care more about religious deliverance from suffering ($\bar{\lambda}_H K$ increases), it is more likely that the location with a higher economic return has a lower economic outcome, $Y(a^*(\bar{\psi}, \tilde{a}^h)) < Y(a^*(\underline{\psi}, \hat{a}^h))$, or a reversal of economic fortune.

In Proposition 4, it is the religious motivation that drives the reversal of economic fortune. Compared to the Western location, the Eastern location had a higher material reward for action. Yet this material motivation must compete with the religious motivation in the opposite direction. When $\bar{\lambda}_H K$ increases, individuals care more about religious deliverance from suffering. More painful suffering presses individuals to obey salvation religions on the margin, an obedience that will discredit these religions. Driven by this more intense search for "salvation," individuals move more towards mystic escape in the Eastern location and more towards activism in the Western location, which further reverses economic fortune.

Heterogeneity in religious motivation. In the above thought experiment, religious motivation is the same in both locations. This is conceptually clean as it does not impose asymmetry in religious motivation. In many historical episodes, however, religious motivation becomes stronger in only one location, say the Western one. Proposition 8 in Appendix A considers this scenario, which obtains similar comparative statics as Proposition 4. The stronger religious motivation in the Western location still induces a higher probability that economic fortune is reversed. In particular, the Eastern location may now mix between activism and mysticism, which represents the East more broadly.

The analysis might help understand the actual process of the great economic reversal in its location and timing. Our model predicts that the process might be pioneered by specific locations in the West that were fully dominated by activist religions. Their economic performance was especially likely to surpass the Eastern economy after major historical episodes had significantly intensified religious motivation in these Western locations.

4 Religious tendencies and executive constraints

Another major payoff from modeling salvation religions is to shed light on a perennial puzzle in constitutionalism: how is it possible to constrain executive power? More specifically,

through the ultimate weapon of **political violence**, can't a powerful chief executive always persecute any opponents into submission? We show that an activist religion can transform the experience of being violently persecuted into an even stronger determination to constrain the chief executive. This remarkable transformation is rooted in the unique capacity of activist religions to endow suffering with meaning that revitalizes action.

4.1 Political domination and religion-motivated resistance

We embed our model of religion in a simple political game. A religio-cultural elite may resist the domination of a king, while the king may persecute the elite. For each period $t = 1, 2, ...$, the timeline is as follows.

1. The elite inherits his belief μ_{t-1}. We will specify the structure of μ_{t-1} in detail later.
2. The elite chooses $a_t \geq 0$ to resist domination.
3. The king chooses to dominate at $D_t \geq 0$. Domination succeeds with probability $Q(a_t) \in [0, 1]$, with $Q' < 0$, and $Q'' \geq 0$. When well defined, two boundary conditions are assumed: $\lim_{a\to\infty} Q(a) > 0$ and $\lim_{a\to\infty} Q'(a) = 0$.
4. With probability $\nu(a_t) \in [0, 1]$, a crisis hits the king's throne, which the king values at $R \sim G(\cdot)$; the king can retain his throne by persecuting the elite at a cost $\kappa(a_t)$, with $\nu' > 0$ and $\kappa' > 0$. If a king loses his throne, he is replaced with another identical king.
5. The elite updates his belief, obtaining μ_t.

The king's problem is simple. At Stage 4, the king persecutes the elite if and only if $R \geq \kappa(a_t)$. So the king's problem at Stage 3 is:

$$\max_{D_t \geq 0} \underbrace{D_t Q(a_t) - \zeta C_k(D_t)}_{\text{payoff from domination}} +$$

$$\underbrace{[1 - \nu(a_t)]\mathbb{E}[R] + \nu(a_t)\Big\{1 - G\big(\kappa(a_t)\big)\Big\}\Big\{\mathbb{E}[R|R \geq \kappa(a_t)] - \kappa(a_t)\Big\}}_{\text{anticipated payoff from persecution}}. \tag{12}$$

In the king's payoff, $D_t Q(a_t)$ is the expected benefit from domination, and $\zeta C_k(D_t)$ is the cost of domination, where $\zeta > 0$ is a parameter, with $C_k' > 0$ and $C_k'' > 0$. The second line of Equation 12 is the payoff from persecution. This payoff comes from the effort to retain the throne (R). With probability $1 - \nu(a_t)$, the king is able to retain his throne without persecution. With probability $\nu(a_t)$, persecution becomes necessary, and the king does so with probability $1 - G\big(\kappa(a_t)\big)$, obtaining a payoff of $\mathbb{E}[R|R \geq \kappa(a_t)] - \kappa(a_t)$.

To ensure that the king's optimal domination is interior, assume that $C_k'(0) < \lim_{a\to\infty} Q(a)/\zeta$. The king's domination D_t^* is thus a decreasing function of a_t, the resistance of the elite:

$$D_t^* = C_k'^{-1}\big(\frac{Q(a_t)}{\zeta}\big) \equiv D(a_t), \text{ with } D' < 0.$$

Now we focus our attention on the religio-cultural elite.

4.2 Activist religions constrain domination

As in the baseline model in Section 2, the religio-cultural elite considers two worldviews. The probability that the elite suffers from political persecution is:

$$\begin{cases} \bar{\lambda}_H \mathbf{1}\{a_t \neq a^h\} & \text{under the religious worldview} \\ \lambda_M & \text{under the naturalist worldview} \end{cases}.$$

At the start of each period t, the elite believes that his religion is true with probability μ_{t-1}. At the end of the period t, the religio-cultural elite updates his belief in the same manner as in the baseline model.

The payoff to the elite is very similar to the baseline model payoff:

$$\max_{a_t\geq 0} \underbrace{w_s[-D(a_t)Q(a_t) - \eta C_e(a_t)]}_{\text{secular payoff from resistance}} \underbrace{+H(a_t - a^h; \mu_{t-1})}_{\text{religious guilt}} \underbrace{-[\mu_{t-1}\bar{\lambda}_H \mathbf{1}\{a_t \neq a^h\} + (1-\mu_{t-1})\lambda_M]K}_{\text{the expected cost of suffering from persecution}}. \tag{13}$$

1. The secular payoff: The term $D(a_t)Q(a_t) > 0$ is the expected rent that the elite surrenders to the king. The surrendered rent $D(a_t)Q(a_t)$ decreases with a_t, the resistance from the elite against domination. The term $\eta C_e(a_t) > 0$ is the secular cost of resistance, where $\eta > 0$ is a parameter, with $C_e' > 0$, $C_e'' > 0$, and $\lim_{a\to\infty} C_e' \to \infty$. The parameter $w_s > 0$ is the weight on the secular payoff.

2. The remaining two components are the same as the baseline model.

To ensure that there is a unique optimal resistance, assume $D''(a_t) \geq 0$ so that the marginal benefit from resistance in deterring domination *decreases* with the level of resistance.[9] We can then define the corresponding secular optimal resistance:

$$a^m = \arg\max_a [-D(a)Q(a) - \eta C_e(a)].$$

[9]Recall that $D' < 0$, so $D'' \geq 0$ implies a *decreasing* marginal benefit from resistance, a reasonable assumption.

Proposition 5. *1. For $a^h \in [a^m, \infty)$, there exists a unique cutoff $\bar{a}(\bar{\lambda}_H K) \in (a^m, \infty)$, such that a religion is stable if and only if $a^h \geq \bar{a}(\bar{\lambda}_H K)$.*

Suppose that w_s is large enough and η is small enough. For $a^h \in [0, a^m]$, there exists a unique cutoff $\underline{a}(\bar{\lambda}_H K) \in (0, a^m)$, such that a religion is stable if and only if $a^h \leq \underline{a}(\bar{\lambda}_H K)$.

2. Consider $a^h > \bar{a}(\bar{\lambda}_H K)$. Whenever the king persecutes the religio-cultural elite in a period t, the elite resists the king even more fiercely in the next period at $a^*_{t+1} > a^*_t$, forcing the king to choose a weaker domination in the next period at $D^*_{t+1} < D^*_t$.

The first part of Proposition 5 simply replicates Proposition 1 for the elite in the political game. We highlight the cutoffs as functions of $\bar{\lambda}_H K$, persecution on a disobeying elite as expected by the religious worldview, which is useful for Section 4.3 on Chinese religions. The upper cutoff $\bar{a}(\bar{\lambda}_H K) \in (a^m, \infty)$ always exists, so activist religions may always guide political action. The lower cutoff $\underline{a}(\bar{\lambda}_H K) \in (0, a^m)$ requires extra conditions because the action space $[0, \infty)$ is bounded from below, while the model under political action does not satisfy all regularity conditions in the baseline model. Intuitively, for zero resistance ($a_t = 0$) to be beyond human capability, the elite should want to resist political domination purely for its material payoff; resistance produces high material payoff when η is low and w_s is high.

The king's stark dilemma under activist religions The second part of Proposition 5 captures the central dilemma of a king who faces an "activist" religio-cultural elite ($a^h > \bar{a}(\bar{\lambda}_H K)$). If the king does not persecute such an elite, the king risks losing his throne immediately. But if the king does persecute the elite, the elite will resist a tyrannical king even more enthusiastically in the future, forcing the king to reduce future domination. The elite is so resolute because of his religious belief, that the elite suffers from political persecution because he has not done *enough* in resisting the king as an active instrument of God. Persecution therefore fortifies the elite's faith in the activist religion, propelling the elite to forever stand firm in constraining the king. This "titanic and inflexible will" (Fukuyama (2011)) of Western religious figures in their struggle against kings is extremely well documented by historians. Appendix D elaborates on the two paradigmatic cases of the great Hebrew prophets in Jewish kingdoms and the religious figures in Western Europe of the Middle Ages.

4.3 Confucianism, Taoism, and the great political divergence

In Imperial China, the dominant religio-cultural elites are the Mandarins, who are the scholar-officials in the Chinese imperial bureaucracy. Mandarins embody the central leit-

motif of Chinese culture, the "complementarity" between the world-adjusting Confucianism and the mystic Taoism (*Rudao Hubu*, see Li (2010)). To model this, the belief system μ_{t-1} is a vector of beliefs in two religious worldviews:

$$\mu_{t-1} = (\mu^c_{t-1}, \mu^d_{t-1}),$$

where $\mu^c_{t-1} \in [0,1)$ is the belief in Confucianism, $\mu^d_{t-1} \in [0,1)$ is the belief in Taoism, and $(1 - \mu^c_{t-1} - \mu^d_{t-1}) \in [0,1)$ is the belief in the naturalist worldview.

Under the three worldviews, the probability of suffering from political persecution in the game of Section 4.1 is:

$$\begin{cases} \bar{\lambda}_C \mathbf{1}\{a_t \neq a^c\} & \text{under Confucianism;} \\ \bar{\lambda}_D \mathbf{1}\{a_t \neq a^d\} & \text{under Taoism;} \\ \lambda_M & \text{under the naturalist worldview.} \end{cases}$$

Assume that $0 < \lambda_M < \bar{\lambda}_D < \bar{\lambda}_C < 1$. We will soon impose restrictions on a^c and a^d to capture Confucianism as a religion of world adjustment and Taoism as a mystic religion. Before that, we specify the payoff of a Chinese Mandarin:

$$\max_{a_t \geq 0} \underbrace{w_s[-D(a_t)Q(a_t) - \eta C_e(a_t)]}_{\text{secular payoff from resistance}} \underbrace{+H(a_t - a^c; \mu^c_{t-1})}_{\text{Confucian guilt}} \underbrace{+H(a_t - a^d; \mu^d_{t-1})}_{\text{Taoist guilt}}$$

$$\underbrace{-[\mu^c_{t-1}\bar{\lambda}_C \mathbf{1}\{a_t \neq a^c\} + \mu^d_{t-1}\bar{\lambda}_D \mathbf{1}\{a_t \neq a^d\} + (1 - \mu^c_{t-1} - \mu^d_{t-1})\lambda_M]K.}_{\text{the expected cost of suffering from persecution}}$$

The payoff extends from the payoff 13, only adjusting for the guilt component and the suffering from persecution to incorporate two religions.

Recall that $-D(a_t)Q(a_t) - \eta C_e(a_t)$ is the secular payoff from resistance. The next assumption reflects that Confucianism is world-adjusting and Taoism is mystic.

Assumption 1. *1.* $-D(a^c)Q(a^c) - \eta C_e(a^c) > -D(a^d)Q(a^d) - \eta C_e(a^d)$.

2. $a^d < \underline{a}(\bar{\lambda}_C K)$, $a^m < a^c < \bar{a}(\bar{\lambda}_C K)$ *where* $\underline{a}(\bar{\lambda}_C K)$ *is as defined in Proposition* 5 *by setting* $\bar{\lambda}_H = \bar{\lambda}_C$.

The first part states that the secular payoff is higher under the Confucian commandment a^c than under the Taoist commandment a^d, reflecting Confucianism's closer alignment with secular motives compared to Taoism.

The second part further states that the Confucian commandment a^c is between a^m and the cutoff $\bar{a}(\bar{\lambda}_C K) > a^m$, which is specified in Proposition 5. Therefore, as a world-adjusting

religion, there is a small distance between the Confucian commandment a^c and the secular optimal action a^m. In the Chinese tradition, Confucianism is also considered "active," therefore $a^c > a^m$ (Tu (1986)). Only a comparison with Judeo-Christianity reveals that Confucianism is not active *enough*, adapting to the world but not revolutionizing the world.

By contrast, the mystic Taoism promises that permanent equanimity flows from "non-action" (*wu wei*), attaining freedom through detachment from the profane world. This is captured by the assumption that $a^d < \underline{a}(\bar{\lambda}_C K)$, which implies that $a^d < \underline{a}(\bar{\lambda}_D K)$.

We also impose a regularity assumption on the guilt function to simplify our analysis.

Assumption 2. *There exists a $\mu_H \in (0,1)$, such that for all $\mu \leq \mu_H$ and $x \in \mathbb{R}$, $H(x;\mu) = 0$. For $\mu \geq \mu_H$, the regularity conditions in Section 2.1.2 remain.*

The assumption says that if a religious worldview is sufficiently unlikely, the individual does not feel internal guilt of acting against its commandment. We are now ready to state the main proposition that formalizes the dynamic "complementarity" between Confucianism and Taoism, along with the serious political consequences.

Proposition 6. *Suppose that the Mandarin's initial beliefs satisfy*

$$\mu_0^c \geq \mu_H, \qquad 0 < \mu_0^d \leq \mu_0^c \mu_H, \qquad 1 - \mu_0^c - \mu_0^d > 0.$$

Then there exist finite integers $I \geq 0$ and $J > I + 1$ such that:

1. *The Mandarin first obeys the Confucian commandment after I realized persecutions. Each of the first I persecutions implies $a_{t+1}^* > a_t^*$ and $D_{t+1}^* < D_t^*$.*

2. *At the $(I+1)$-st persecution, the Mandarin is obeying the Confucian commandment. Immediately afterward, his posterior belief in Confucianism is zero.*

3. *After more than J persecutions, each new persecution implies $a_{t+1}^* < a_t^*$ and $D_{t+1}^* > D_t^*$.*

A comparative analysis of religio-cultural elites and political domination Proposition 6 is illustrated in Figure 5. The upper and middle panels compare a Chinese Mandarin (in orange) with an activist Western "prophet" (in blue) of Proposition 5. The upper panel shows the beliefs of these religio-cultural elites, and the middle panel shows their resistance against domination, both over how many times the two elites have suffered from persecution. Notice that in the short run, the belief and action of the Chinese Mandarin closely track those of the activist prophet. Upon being persecuted, both the Chinese Mandarin and the Western prophet intensify their resistance against domination. But the Confucian commandment is ultimately within human capability. The Mandarin will fulfill the Confucian commandment,

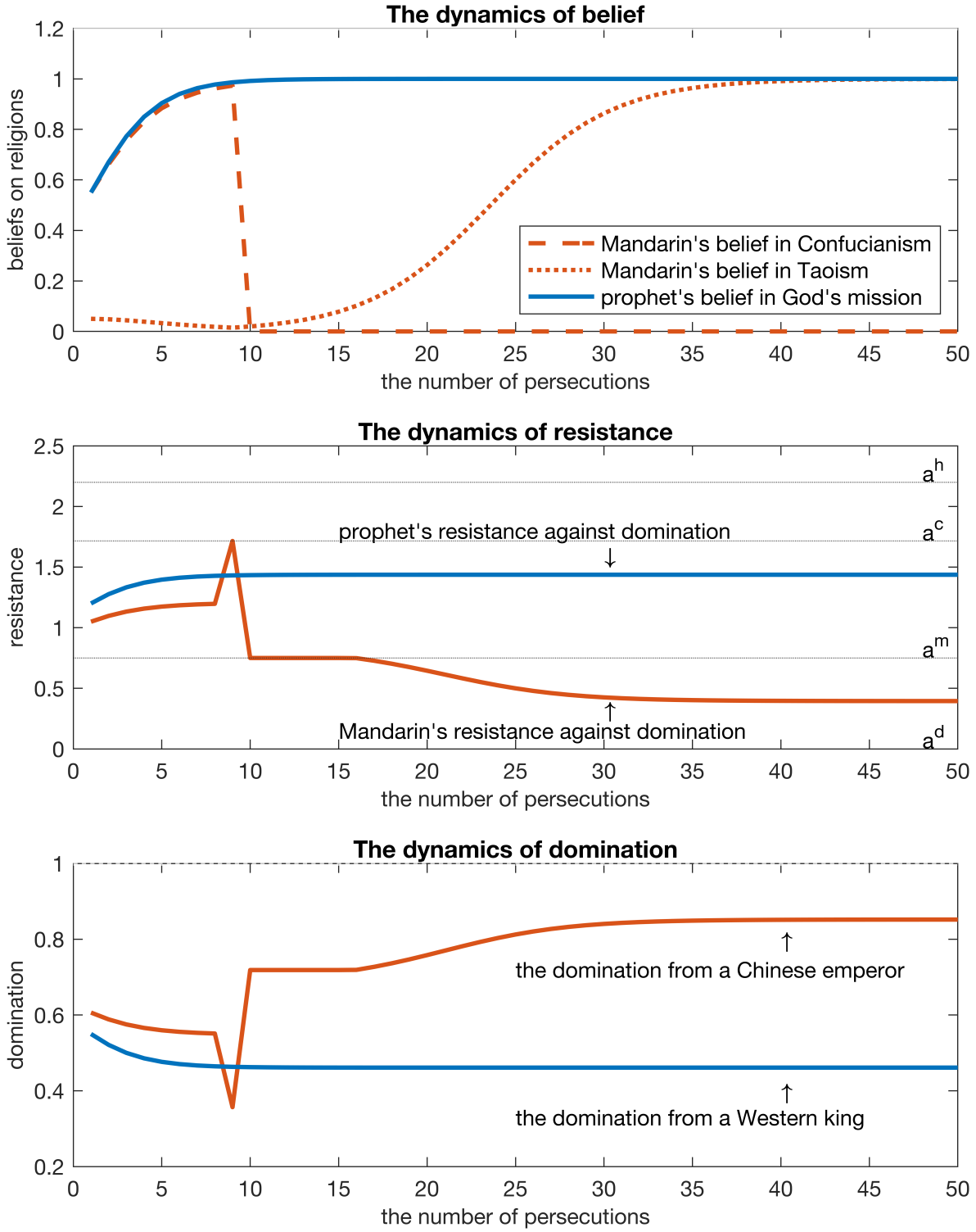


Figure 5: A comparison between a Western "prophet" and a Chinese Mandarin, as well as the domination from a Western king and a Chinese emperor

which will eventually disenchant Confucianism. The disheartened Mandarin then gradually accepts Taoism, reasoning that he suffers from political persecution because he did not escape far enough from the profane world of political action. Under the guidance of Taoist "non-action," the Mandarin escapes from the Confucian duty to discipline the emperor, in sharp contrast to the unwaveringly strong resistance from the Western prophet.

The lower panel in Figure 5 traces out the dynamics of political domination from the Chinese emperor versus the Western king. In China, the emperor might be "accountable" for a while, but such accountability is transient. The Chinese emperor eventually achieved much stronger domination than the Western king, induced by the ultimate collapse of resistance from the Mandarin versus the unwavering activism of the Western prophet.[10]

4.4 Religious legitimation and church power

Our analysis so far focuses on how religion may succeed or fail in providing persistent motivation to resist political power. At the same time, many religions also provide effective legitimation of political power (Rubin (2017); Bentzen and Gokmen (2023); Bisin et al. (2024); Greif and Rubin (2024)). Building on our model of salvation religions, we investigate how religion may legitimize political power in another related paper (Li and Wang (2026)). We show that a strict religion, either activist or mystic, is the most effective legitimizer of political power. Under these religions, the sinfulness of a king's subjects is a sufficient explanation for disasters, which protects the political legitimacy of the king. By contrast, world-adjusting religions are less capable of legitimizing political power because the moral goodness of the subjects will shift blame for disasters to the king, who may have lost the "mandate of heaven."

Since a strict religion is effective at legitimizing the king, the king has to concede substantial economic surplus to the church. Thus, strict churches are highly organized partly because they can extract significant material resources from the state to support their large personnel. This helps understand the seeming paradox that mystic religions, while least able to motivate worldly action, can still produce highly organized churches (Carvalho and Sacks (2021)).

While religious activism and mysticism are equally powerful legitimizers, it is still an activist religion that constrains the king most effectively. Like mystic religions, activist religions can extract substantial surplus and other concessions from the king. But unlike mystic religions, activist religions might further motivate their religious personnel to actively deter a king from transgression, as our previous analysis has shown.

[10] Xue (2021) presents empirical evidence that persecution of Chinese literati leads to "political apathy."

5 Religious tendencies and competition for status

The baseline model begs a further question. How could Confucianism be so dominant in Imperial China, or for any other world-adjusting religions? This section investigates an alternative route to protect the "salvation" promise. A world-adjusting religion can establish a competitive contest that certifies an "elite" status, promising that world adjustment achieves "salvation," but only for certified elites. Salvation through such an elite status is credible among "commoners" who are unable to directly falsify the promise. But these status contests are less likely to emerge under a salvation religion because their strict commandments are already sufficient to protect their credibility.

There is indeed a close connection between world affirmation and status contests:

- Chinese society long revolved around the Confucian civil service examination, which conferred the elite status of Mandarins (Bai and Jia (2016); Chen et al. (2020)).
- The competition over socioeconomic status is fierce in the "modern world," whose dominant culture affirms humanity itself "as the source of truth and moral conduct" (pp.14, Troeltsch (2017); Turchin (2023)).

In our model, these status contests sustain the credibility that world affirmation can free one from all troubles. Equally important, we will show that only a culture of world affirmation can induce status contests at such extravagance.

To formalize these ideas, we extend the baseline model by introducing two social classes, elites and commoners. At the start of each period t, there is a measure $m_E > 0$ of elites and a measure 1 of commoners. At each period t, a $q > 0$ fraction of elites retire and exit the game, leaving $m_E q < 1$ elite positions available for commoners to compete. To keep the total population constant at $1 + m_E$, assume that at each period, $m_E q$ new commoners are born and enter the game.

The commoners can compete in a contest that grants elite status. Specifically, a commoner indexed by $i \in [0, 1]$ chooses an effort e_t^i in the contest at a cost of $\Gamma(e_t^i)$. The probability that the commoner succeeds in the contest is:

$$\Pi(e_t^i; e_t^{-i}),$$

where

$$e_t^{-i} = \int_{j \in [0,1] \setminus \{i\}} e_t^j dj$$

is the average level of effort of other commoners. Assume that $\Pi(e_t^i; e_t^{-i})$ increases with e_t^i and decreases with e_t^{-i}, reflecting the tournament nature of such a contest. For interior

solutions, also assume $\frac{\partial \Pi(0;0)}{\partial e_t^i} > 0$.

The naturalist or mechanical worldview is the same as in the baseline model. The (modified) religious worldview is:

$$\lambda_H = \begin{cases} 0 & \text{if and only if } a_t^i = a^h \text{ as an elite} \\ \bar{\lambda}_H & \text{otherwise} \end{cases}. \tag{14}$$

That is, one is free from troubles if one obeys the commandment a^h *as an elite.* Elites are privileged here because they have an opportunity to attain "salvation."

After the contest, the payoff to an individual i is

$$\psi F(a_t^i) - C(a_t^i) + H(a_t^i - a^h; \mu_{t-1}) - [\mu_{t-1}^i \bar{\lambda}_H \cdot \underbrace{\mathbf{1}\{a_t^i \neq a^h \text{or } i \text{ is not an elite}\}}_{\text{the only difference in payoff}} + (1 - \mu_{t-1}^i)\lambda_M]K.$$

The only difference from the payoff in the baseline model is highlighted (except the superscript i in the action a_t^i and the belief μ_{t-1}^i). To reiterate, the only difference is that "salvation" is always unattainable to an individual *as a commoner.*

We simplify the analysis by looking at a situation where all commoners are identical and experience common shocks. We can then only track the belief and the strategies of a representative commoner. So we ignore the superscript i for the representative commoner, and his belief is simply μ_{t-1}. The timeline of the game for each period t is as follows.

1. All individuals inherit their beliefs from the last period.
2. A $m_E q$ measure of elites retire and exit the game.
3. Commoners choose efforts e_t^i for all $i \in [0, 1]$. This determines the probability of contest success and the new cohort of elites. The successful commoners are replaced with newborns, who inherit the belief μ_{t-1} from existing commoners.
4. All individuals choose their actions.
5. All individuals, including the elites and commoners, may suffer.
6. All individuals update their beliefs. Specifically, commoners obtain μ_t.

We impose standard regularity conditions for this simple game of status contests. The second-order condition for e_t^i requires a decreasing marginal return to one's own effort ($\frac{\partial^2 \Pi}{\partial (e_t^i)^2} < 0$) and an increasing marginal cost from effort ($\Gamma' > 0$, $\Gamma'(0) = 0$, and $\Gamma'' > 0$). To ensure the uniqueness of the equilibrium, we also assume the following condition.

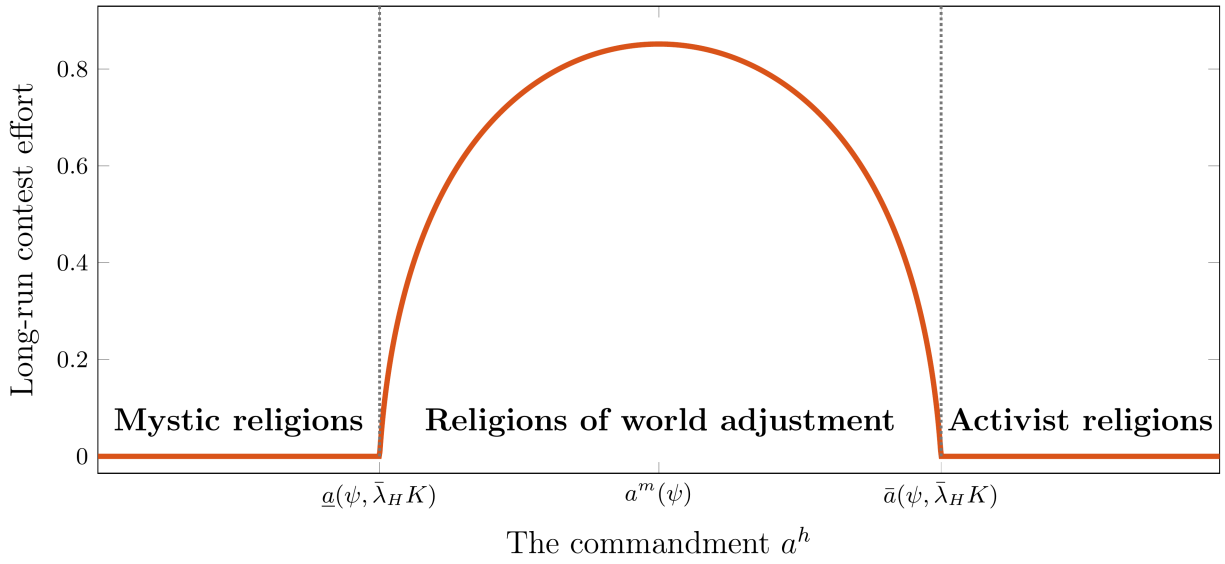


Figure 6: Religious tendencies and status competition

Assumption 3.

$$\frac{\partial^2 \Pi}{\partial e_t^i \partial e_t^{-i}} < 0.$$

This assumption guarantees a regular best response function: a commoner reduces his effort if other commoners increase their effort. Without this assumption, there might be a strategic complementarity among the efforts of commoners, a complementarity that could generate multiple equilibria in efforts.

Proposition 7. *There is a unique symmetric equilibrium in the commoners' efforts in the status contest, denoted as* e_t^*.

1. *Suppose that* a^h *is a world-adjusting religion:* $a^h \in \left(\underline{a}(\psi, \bar{\lambda}_H K), \bar{a}(\psi, \bar{\lambda}_H K)\right)$ *as defined in Proposition 2.*

   (a) *Every time the representative commoner suffers, the more believing commoner exerts a higher effort in the contest.*

   (b) *The long-run effort* $\lim_{t\to\infty} e_t^*$ *is strictly positive after a sufficiently large number of sufferings. The long-run effort strictly increases with* a^h *if* $a^h < a^m(\psi)$ *and strictly decreases with* a^h *if* $a^h > a^m(\psi)$.

2. *Suppose that* a^h *is a salvation religion:* $a^h \in [0, \infty) \backslash \left(\underline{a}(\psi, \bar{\lambda}_H K), \bar{a}(\psi, \bar{\lambda}_H K)\right)$ *as in Proposition 1. Then commoners always exert zero efforts in the contest:* $e_t^* = 0$.

In Proposition 7, Part 1(a) is straightforward. A suffering commoner concludes that he suffers because he is not an elite, since elites can attain "salvation" through the easy action

of world adjustment. The more believing commoner thus exerts more effort in the contest to become an elite.

The rest of the proposition is illustrated in Figure 6, where the horizontal axis is the religious commandment a^h, and the vertical axis is the long-run equilibrium effort in the contest after an arbitrarily large number of sufferings. Figure 6 and Proposition 7 illustrate the central insight of this section, the connection between world affirmation and status competition. While the long-run effort in the status competition is strictly positive for a world adjusting religion, commoners do not contest the status at all under a salvation religion, either mystic or activist. Why this contrast?

World affirmation supports status competition Under a world adjusting religion, commoners believe that elites have attained the commandment that frees them from all troubles. Therefore, there is a striking contrast between the elites in permanent bliss and commoners in permanent danger of suffering, *as believed by commoners*. Commoners can sustain this belief because they have never experienced the life of elites. This belief then drives the fierce competition into the elite status. The competition for the elite status is especially extravagant when the religion strongly affirms the "human nature" (a very small distance between a^h and $a^m(\psi)$), so that the life of an elite is believed to be especially blissful: the elite attains "salvation" simply by fulfilling his humanly desires.

Salvation religions discourage status competition Status competition is not necessary for salvation religions, which can fully protect their credibility by their superhuman commandments. So salvation religions do not need to set up a status competition in the first place.[11] What if a salvation religion still does so? Part 2 of Proposition 7 investigates this situation. Commoners never contest the elite status because they realize that elites, even though endowed with the opportunity to attain "salvation," still cannot do so because no humans can fulfill the superhuman commandment. Suffering is driven by the constitutive imperfection in "human nature" that transcends social classes (pp.330, Weber (2013)).

The theoretical analysis is summarized in Figure 6, showing an alignment between status competition and world adjusting religions. Appendix E discusses in detail the implications of the theoretical analysis, focusing on the civil service examination in Imperial China, as well as the prominence of status competition in the modern world. Appendix E also shows that our results are robust under modified setups. Commoners may compete for the elite status for

[11] This can be easily formalized by introducing a "designer" for any religion a^h at $t = 0$, before the game starts at $t = 1$. A designer obtains $U(x)$ from a follower with a belief at $x \in [0, 1]$, with $U'(\cdot) > 0$. Suppose that $U(1)$ is larger than a strictly positive fixed cost to set up a status competition. It is obvious that a designer introduces a status competition if and only if the religion is world-adjusting.

both the opportunity of "salvation" and an extra material benefit $B > 0$. Commoners may also feel no guilt from disobedience before they succeed in the status competition. In either case, there is still a strong alignment between status competition and world adjustment.

6 Conclusion

Our analysis of religions is motivated by the central principle that only the unattainable can persistently enchant, the principle that drives individuals to move away and transcend their "human nature." This self-transcendence has been identified as essential to religion.

> [R]eligion is the capacity of human organism to transcend its biological nature through the construction of objective, morally binding, all embracing universes of meaning (pp.176, Berger (1967)).

In our model, the religious worldview can achieve objectivity as full belief and morally binding force as a strong guidance of human action, but only when it demands the transcendence beyond human will power to protect its plausibility (Proposition 1). Pioneered by the club and social approach to church strictness (Iannaccone (1992); Carvalho (2013)), this view on religion is also central to our belief-based analysis of salvation religions. Among others, the paradoxical nature of self-transcendence sheds light on the great reversal of economic fortune between the East and the West. In our model, salvation religions can reverse economic fortune because self-transcendence demands world renunciation in a productive economy and world mastery in a backward economy (Proposition 3 and Proposition 4).

We also show that salvation religions as self-transcendence may shed light on how it is possible to constrain political power (Section 4). Violence as the ultimate weapon of political power can be neutralized by an unwavering conviction in the value of resisting tyranny against self interests, a conviction that is paradoxically driven by a selfish human nature.

The principle of unattainability can also bridge the analysis of salvation religions and "quasi-religions," especially the quest for socioeconomic status that is extremely difficult to attain (Tillich (1963); Tillich (2009)). Section 5 shows that the principle sheds light on why these "quasi-religions" are especially dominant in a culture of world affirmation, where the influence of salvation religions has declined. We therefore follow the Tillichian program (Tillich (2009)) to analyze religions and "quasi-religions" under a unified principle, which also helps reveal their fundamental differences.

Transcendence of selfish nature may generate many other unintended consequences. Mechanisms that could preserve peace among self-interested individuals, such as the "com-

mercial peace" (Hirschman (2013)), began to crack. Novel forms of conflicts could emerge, conflicts that were unthinkable among purely self-interested individuals (Berman and Laitin (2008); Carvalho and Sacks (2024)). Further investigation is warranted for the many paradoxical relations between actions and ideologies that are sustained by the "unattainability" condition.

References

Abdulkadiroğlu, Atila, Joshua Angrist, and Parag Pathak. 2014. The elite illusion: Achievement effects at Boston and New York exam schools. *Econometrica* **82**: 137–196.

Abu-Lughod, Janet L. 1989. *Before European Hegemony: The World System AD 1250-1350*. Oxford University Press.

Acemoglu, Daron, Simon Johnson, and James A Robinson. 2002. Reversal of fortune: Geography and institutions in the making of the modern world income distribution. *The Quarterly Journal of Economics* **117**: 1231–1294.

Acemoglu, Daron, and James A. Robinson. 2023. Weak, despotic, or inclusive? how state type emerges from state versus civil society competition. *American Political Science Review* **117**: 407–420.

Aimone, Jason A, Laurence R Iannaccone, Michael D Makowsky, and Jared Rubin. 2013. Endogenous group formation via unproductive costs. *Review of Economic Studies* **80**: 1215–1236.

Akerlof, Robert. 2017. The importance of legitimacy. *The World Bank Economic Review* **30**: S157–S165.

Alesina, Alberto, and Paola Giuliano. 2015. Culture and institutions. *Journal of Economic Literature* **53**: 898–944.

Augenblick, Ned, Jesse M Cunha, Ernesto Dal Bó, and Justin M Rao. 2016. The economics of faith: using an apocalyptic prophecy to elicit religious beliefs in the field. *Journal of Public Economics* **141**: 38–49.

Bai, Ying, and Ruixue Jia. 2016. Elite recruitment and political stability: The impact of the abolition of China's civil service exam. *Econometrica* **84**: 677–733.

Barros, Pedro Pita, and Nuno Garoupa. 2002. An economic theory of church strictness. *The Economic Journal* **112**: 559–576.

Baumeister, Roy F, Ellen Bratslavsky, Catrin Finkenauer, and Kathleen D Vohs. 2001. Bad is stronger than good. *Review of General Psychology* **5**: 323–370.

Bazzi, Samuel, Masyhur Hilmy, and Benjamin Marx. 2026. Religion, education, and the state. *The Review of Economic Studies* **93**: 1494–1535.

Bazzi, Samuel, Gabriel Koehler-Derrick, and Benjamin Marx. 2020. The institutional foundations of religious politics: Evidence from Indonesia. *The Quarterly Journal of Economics* **135**: 845–911.

Becker, Sascha O., and Steven Pfaff. 2023. Church, state, and historical political economy. In Jenkins, Jeffery A., and Jared Rubin (eds.) *The Oxford Handbook of Historical Political Economy*. Oxford University Press, 925–944.

Becker, Sascha O., Jared Rubin, and Ludger Wößmann. 2024. Religion and Growth. *Journal of Economic Literature* **62**: 1094–1142.

Bellah, Robert N. 1999. Max Weber and world-denying love: A look at the historical sociology of religion. *Journal of the American Academy of Religion* **67**: 277–304.

Belloc, Marianna, Francesco Drago, and Roberto Galbiati. 2016. Earthquakes, religion, and transition to self-government in Italian cities. *The Quarterly Journal of Economics* **131**: 1875–1926.

Bénabou, Roland, and Jean Tirole. 2006. Belief in a just world and redistributive politics. *The Quarterly Journal of Economics* **121**: 699–746.

Bénabou, Roland, and Jean Tirole. 2011. Identity, morals, and taboos: Beliefs as assets. *The Quarterly Journal of Economics* **126**: 805–855.

Bentzen, Jeanet Sinding, and Gunes Gokmen. 2023. The power of religion. *Journal of Economic Growth* **28**: 45–78. ISSN 1381-4338.

Bentzen, Jeanet Sinding. 2019. Acts of God? Religiosity and natural disasters across subnational world districts. *The Economic Journal* **129**: 2295–2321.

Bentzen, Jeanet Sinding. 2021. In crisis, we pray: Religiosity and the covid-19 pandemic. *Journal of Economic Behavior & Organization* **192**: 541–583.

Bentzen, Jeanet Sinding, and Eric R Force. 2023. Did seismic activity lead to the rise of religions? In *The Economics of Religion*. World Scientific, 63–95.

Berger, Peter L. 1967. *The Sacred Canopy: Elements of a Sociological Theory of Religion*. Knopf Doubleday Publishing Group.

Bergeron, Augustin, Jean-Paul Carvalho, Joseph Henrich, Nathan Nunn, and Jonathan L. Weigel. 2026. Zero-sum environments, the evolution of effort-suppressing beliefs, and economic development. *The Review of Economic Studies* Forthcoming.

Berman, Eli. 2000. Sect, subsidy, and sacrifice: an economist's view of ultra-Orthodox Jews. *The Quarterly Journal of Economics* **115**: 905–953.

Berman, Eli, and David D Laitin. 2008. Religion, terrorism and public goods: Testing the club model. *Journal of Public Economics* **92**: 1942–1967.

Binzel, Christine, and Jean-Paul Carvalho. 2017. Education, social mobility and religious movements: The Islamic revival in Egypt. *The Economic Journal* **127**: 2553–2580.

Bisin, Alberto, Jared Rubin, Avner Seror, and Thierry Verdier. 2024. Culture, institutions and the long divergence. *Journal of Economic Growth* **29**: 1–40.

Bisin, Alberto, Giorgio Topa, and Thierry Verdier. 2004. Religious intermarriage and socialization in the United States. *Journal of Political Economy* **112**: 615–664.

Bohren, J Aislinn, and Daniel N Hauser. 2025. Misspecified models in learning and games. *Annual Review of Economics* **17**: 427–451.

Butinda, Lewis Dunia, Aimable Amani Lameke, Nathan Nunn, Max Posch, and Raúl Sánchez de la Sierra. 2023. On the Importance of African Traditional Religion for Economic Behavior. NBER Working Paper 31430, National Bureau of Economic Research.

Carvalho, Jean-Paul. 2013. Veiling. *The Quarterly Journal of Economics* **128**: 337–370.

Carvalho, Jean-Paul. 2020. Sacrifice and sorting in clubs. *Forum for Social Economics* **49**: 357–369.

Carvalho, Jean-Paul, and Michael Sacks. 2021. The economics of religious communities. *Journal of Public Economics* **201**: 104481.

Carvalho, Jean-Paul, and Michael Sacks. 2024. Radicalisation. *The Economic Journal* **134**: 1019–1068.

Chaney, Eric. 2013. Revolt on the Nile: Economic shocks, religion, and political power. *Econometrica* **81**: 2033–2053.

Chao, Shin-yi. 2003. Daoist examinations and Daoist schools during the Northern Song dynasty. *Journal of Chinese Religions* **31**: 1–37.

Chen, Ting, James Kai-sing Kung, and Chicheng Ma. 2020. Long live Keju! The persistent effects of China's civil examination system. *The Economic Journal* **130**: 2030–2064.

Della Lena, Sebastiano, Elena Manzoni, and Fabrizio Panebianco. 2023. On the transmission of guilt aversion and the evolution of trust. *Games and Economic Behavior* **142**: 765–793.

Egorov, Georgy, and Konstantin Sonin. 2024. The political economics of non-democracy. *Journal of Economic Literature* **62**: 594–636.

Elvin, Mark. 1973. *The Pattern of the Chinese Past: A Social and Economic Interpretation*. Stanford University Press.

Espín-Sánchez, José-Antonio, Salvador Gil-Guirado, and Nicholas Ryan. 2026. Praying for rain. *The Quarterly Journal of Economics* **141**: 2363–2422.

Fearon, James D. 2011. Self-enforcing democracy. *The Quarterly Journal of Economics* **126**: 1661–1708.

Fernández-Villaverde, Jesús, Mark Koyama, Youhong Lin, and Tuan-Hwee Sng. 2023. The fractured-land hypothesis. *The Quarterly Journal of Economics* **138**: 1173–1231.

Finer, Samuel Edward. 1997. *The History of Government from the Earliest Times Volume I: Ancient Monarchies and Empires*, volume 1. Oxford University Press.

Francois, Patrick, Francesco Trebbi, and Kairong Xiao. 2023. Factions in nondemocracies: Theory and evidence from the Chinese Communist Party. *Econometrica* **91**: 565–603.

Fukuyama, Francis. 2011. *The Origins of Political Order: From Prehuman Times to the French Revolution*. Farrar, Straus and Giroux.

Gernet, Jacques. 1995. *Buddhism in Chinese Society: An Economic History from the Fifth to the Tenth Centuries*. Columbia University Press.

Giuliano, Paola, and Nathan Nunn. 2021. Understanding cultural persistence and change. *The Review of Economic Studies* **88**: 1541–1581.

Greif, Avner. 2006. *Institutions and the Path to the Modern Economy: Lessons from Medieval Trade*. Cambridge University Press.

Greif, Avner, Joel Mokyr, and Guido Tabellini. 2025. *Two Paths to Prosperity: Culture and Institutions in Europe and China, 1000–2000*. Princeton University Press.

Greif, Avner, and Jared Rubin. 2024. Endogenous political legitimacy: The Tudor roots of England's constitutional governance. *The Journal of Economic History* **84**: 655–689.

Greif, Avner, and Guido Tabellini. 2017. The clan and the corporation: Sustaining cooperation in China and Europe. *Journal of Comparative Economics* **45**: 1–35.

Hayes, Christine. 2012. *Introduction to the Bible*. (The Open Yale Courses). Yale University Press.

Hirschman, Albert O. 2013. *The Passions and the Interests: Political Arguments for Capitalism before Its Triumph*. Princeton University Press.

Hood Jr, Ralph W, Peter C Hill, and Bernard Spilka. 2018. *The Psychology of Religion: An Empirical Approach*. Guilford Publications.

Iannaccone, Laurence R. 1992. Sacrifice and stigma: Reducing free-riding in cults, communes, and other collectives. *Journal of Political Economy* **100**: 271–291.

Jia, Ruixue, and James Kung. 2025. The Culture and Institutions of Confucianism. In Enke, Benjamin, Paola Giuliano, Nathan Nunn, and Leonard Wantchekon (eds.) *Handbook of Culture and Economic Behavior*.

Jia, Ruixue, Gérard Roland, and Yang Xie. 2024. A Theory of Power Structure and Institutional Compatibility: China versus Europe Revisited. *Journal of the European Economic Association* **22**: 1275–1318.

Landes, David S. 2006. Why Europe and the West? Why not China? *Journal of Economic Perspectives* **20**: 3–22.

Le Rossignol, Etienne, Sara Lowes, and Nathan Nunn. 2025. Culture at a Crossroads: Traditional Beliefs in Contemporary Africa. Revision requested at *the Quarterly Journal of Economics*.

Lee, Seung-Hwan. 1992. Was there a concept of rights in Confucian virtue-based morality? *Journal of Chinese Philosophy* **19**: 241–261.

Levy, Gilat, and Ronny Razin. 2012. Religious beliefs, religious participation, and cooperation. *American Economic Journal: Microeconomics* **4**: 121–151.

Li, Weijia, and Shaoda Wang. 2026. Ethical religions and political legitimacy. Working paper.

Li, Zehou. 2010. *The Chinese Aesthetic Tradition*. University of Hawaii Press Honolulu. Translated by Maija Bell Samei.

McBride, Michael. 2008. Religious pluralism and religious participation: A game theoretic analysis. *American Journal of Sociology* **114**: 77–106.

McCleary, Rachel M, and Robert J Barro. 2006. Religion and economy. *Journal of Economic Perspectives* **20**: 49–72.

Mengzi. 2008. *Mengzi: With Selections from Traditional Commentaries*. Hackett Publishing Company. Translated, with an introduction, by Bryan W. Van Norden.

Michalopoulos, Stelios, and Melanie Meng Xue. 2021. Folklore. *The Quarterly Journal of Economics* **136**: 1993–2046.

Mokyr, Joel. 2016. *A Culture of Growth: The Origins of the Modern Economy*. Princeton University Press.

Nietzsche, Friedrich. 1998. *On the Genealogy of Morality*. Hackett Publishing. Translation and Notes by Maudemarie Clark and Alan J. Swensen, Introduction by Maudemarie Clark.

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

Norenzayan, Ara, and Ian G Hansen. 2006. Belief in supernatural agents in the face of death. *Personality and Social Psychology Bulletin* **32**: 174–187.

Nunn, Nathan. 2025. Cultures in crisis? cultural evolution and the potential for global mismatch. 2025 Griffin Economic History Public Lecture. Link.

Nunn, Nathan, and Raúl Sánchez de la Sierra. 2017. Why being wrong can be right: Magical warfare technologies and the persistence of false beliefs. *American Economic Review* **107**: 582–587.

Pomeranz, Kenneth. 2000. *The Great Divergence: China, Europe, and the Making of the Modern World Economy*. Princeton University Press.

Purzycki, Benjamin Grant, Coren Apicella, Quentin D Atkinson, Emma Cohen, Rita Anne McNamara, Aiyana K Willard, Dimitris Xygalatas, Ara Norenzayan, and Joseph Henrich.

2016. Moralistic gods, supernatural punishment and the expansion of human sociality. *Nature* **530**: 327–330.

Rozin, Paul, and Edward B Royzman. 2001. Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review* **5**: 296–320.

Rubin, Jared. 2017. *Rulers, Religion, and Riches: Why the West got rich and the Middle East did not*. Cambridge University Press.

Stasavage, David. 2020. *The Decline and Rise of Democracy: A Global History from Antiquity to Today*. Princeton University Press.

Swedberg, Richard, and Ola Agevall. 2016. *The Max Weber Dictionary: Key Words and Central Concepts*. Stanford University Press.

Tillich, Paul. 1963. *Christianity and the Encounter of the World Religions*. Columbia University Press.

Tillich, Paul. 2008. *The Courage to Be*. Yale University Press. The second edition.

Tillich, Paul. 2009. *Dynamics of Faith*. HarperCollins.

Troeltsch, Ernst. 1931. *The Social Teaching of the Christian Churches*. George Allen & Unwin. Translated by Olive Wyon; 2 vols.

Troeltsch, Ernst. 2017. *Protestantism and Progress: a Historical Study of the Relation of Protestantism to the Modern World*. Routledge.

Tu, Wei-ming. 1986. The Structure and Function of the Confucian Intellectual in Ancient China. In Eisenstadt, Shmuel N. (ed.) *The Origins and Diversity of Axial Age Civilizations*, chapter 15. Albany, NY: State University of New York Press, 321–347.

Tu, Wei-ming. 1993. *Way, Learning, and Politics: Essays on the Confucian Intellectual*. SUNY Press.

Turchin, Peter. 2023. *End times: Elites, Counter-Elites, and the Path of Political Disintegration*. Penguin.

Ullmann, Walter. 1975. *Law and Politics in the Middle Ages: An Introduction to the Sources of Medieval Political Ideas*. Hodder & Stoughton, London.

Vaish, Amrisha, Tobias Grossmann, and Amanda Woodward. 2008. Not all emotions are created equal: The negativity bias in social-emotional development. *Psychological Bulletin* **134**: 383–403.

Wallace-Hadrill, John Michael. 1965. The Via Regia of the Carolingian Age. In Smalley, Beryl (ed.) *Trends in Medieval Political Thought*. Blackwell, Oxford, 22–41.

Weber, Max. 1952. *Ancient Judaism*. The Free Press. Translated by Hans H. Gerth and Don Martindale.

Weber, Max. 1978. *Economy and Society: An Outline of Interpretive Sociology*. University of California Press. Edited by Guenther Roth and Claus Wittich.

Weber, Max. 1992. *The Protestant Ethic and the Spirit of Capitalism*. Routledge. Translated by Talcott Parsons.

Weber, Max. 2004a. Intermediate reflection on the economic ethics of the world religions. In Whimster, Sam (ed.) *The Essential Weber: A Reader*. Routledge, 215–244.

Weber, Max. 2004b. Introduction to the economic ethics of the world religions. In Whimster, Sam (ed.) *The Essential Weber: A Reader*. Routledge, 55–80.

Weber, Max. 2013. *From Max Weber: Essays in Sociology*. Routledge. Translated, edited, and with an introduction by H. H. Gerth and C. Wright Mills.

White, Cindel JM, John Michael Kelly, Azim F Shariff, and Ara Norenzayan. 2019. Supernatural norm enforcement: Thinking about karma and God reduces selfishness among believers. *Journal of Experimental Social Psychology* **84**: 103797.

White, Cindel JM, Aiyana K Willard, Adam Baimel, and Ara Norenzayan. 2021. Cognitive pathways to belief in karma and belief in god. *Cognitive Science* **45**: e12935.

Xue, Melanie Meng. 2021. Autocratic rule and social capital: Evidence from imperial china. Working paper, SSRN 2856803.

Zürcher, Erik. 1959. *The Buddhist Conquest of China: The Spread and Adaptation of Buddhism in Early Medieval China*. Number 11 in Sinica Leidensia. Leiden: E. J. Brill. 2 vols.

A Proofs for propositions in the text

Proposition 1. *1. A religion is stable if and only if its commandment $a^h \geq 0$ is sufficiently far away from the secular optimal action $a^m(\psi)$.*

Formally, there exist a unique cutoff $\underline{a}(\psi, \bar{\lambda}_H K) \in \big(0, a^m(\psi)\big)$ and a unique cutoff $\bar{a}(\psi, \bar{\lambda}_H K) \in \big(a^m(\psi), \infty\big)$, such that a religion a^h is stable if and only if

$$a^h \leq \underline{a}(\psi, \bar{\lambda}_H K) \text{ or } a^h \geq \bar{a}(\psi, \bar{\lambda}_H K). \tag{6}$$

Both cutoffs $\underline{a}(\psi, \bar{\lambda}_H K)$ and $\bar{a}(\psi, \bar{\lambda}_H K)$ are functions of ψ, the material return to action, and $\bar{\lambda}_H K$, suffering from disobedience as expected by the religion.

2. Under a stable religion, the individual's action moves ever closer to the commandment a^h, but the individual never exactly obeys the commandment a^h.

Formally, denote a_t^* as the individual's action at period t. Under a stable religion a^h, for all $t = 1, 2, \ldots$, $|a_{t+1}^* - a^h| \leq |a_t^* - a^h|$, and $a_{t+1}^* \neq a^h$.

Proof. We first establish a few properties of the individual's action. Because

$$\psi F''(a) - C''(a) < 0,$$

the secular payoff is strictly concave. Moreover,

$$\lim_{a \downarrow 0} [\psi F(a) - C(a)] = -\infty \quad \text{and} \quad \lim_{a \to \infty} [\psi F(a) - C(a)] = -\infty.$$

The second limit is true because under concavity, F grow at most linearly, while the marginal cost goes to infinity. Therefore, $a^m(\psi)$ is unique and interior, and

$$\psi F'(a^m(\psi)) - C'(a^m(\psi)) = 0. \tag{15}$$

Fix a belief $\mu_{t-1} > 0$ and a commandment $a^h \neq a^m(\psi)$. Conditional on disobedience, the expected cost of suffering does not change with action. The individual therefore uniquely maximizes

$$\psi F(a_t) - C(a_t) + H(a_t - a^h; \mu_{t-1}).$$

The optimal action is interior and satisfies

$$\psi F'(a_t^*) - C'(a_t^*) + H'(a_t^* - a^h; \mu_{t-1}) = 0. \tag{16}$$

If $0 < a^h < a^m(\psi)$, the derivative of this objective is positive at a^h and negative at $a^m(\psi)$. So strict concavity implies

$$a^h < a_t^* < a^m(\psi). \tag{17}$$

The same inequalities hold for $a^h = 0$: the payoff at zero is negative infinity, while the derivative at $a^m(\psi)$ is negative. Similarly, if $a^h > a^m(\psi)$,

$$a^m(\psi) < a_t^* < a^h. \tag{18}$$

We now compare obedience and disobedience under full belief. For $a^h > 0$ and $a^h \neq a^m(\psi)$, the maximal payoff from disobedience minus the payoff from obedience is

$$\max_{a_t>0} \Big\{\psi F(a_t) - C(a_t) + H(a_t - a^h; 1)\Big\} - \bar{\lambda}_H K - \Big\{\psi F(a^h) - C(a^h)\Big\}. \tag{19}$$

By the envelope theorem and Equation 16, the derivative of Expression 19 with respect to a^h is

$$\Big[\psi F'(a_t^*) - C'(a_t^*)\Big] - \Big[\psi F'(a^h) - C'(a^h)\Big].$$

The secular marginal payoff is strictly decreasing. Equations 17 and 18 therefore imply that Expression 19 is strictly decreasing for $a^h < a^m(\psi)$ and strictly increasing for $a^h > a^m(\psi)$.

As a^h approaches $a^m(\psi)$, Expression 19 approaches $-\bar{\lambda}_H K < 0$. As $a^h \downarrow 0$, choosing $a_t = a^m(\psi)$ while disobeying gives the lower bound

$$\psi F(a^m(\psi)) - C(a^m(\psi)) + H(a^m(\psi) - a^h; 1) - \bar{\lambda}_H K - \Big[\psi F(a^h) - C(a^h)\Big],$$

which approaches infinity because $\psi F(a^h) - C(a^h) \to -\infty$. In particular, obedience at $a^h = 0$ gives a payoff of negative infinity.

Finally, as $a^h \to \infty$, consider the disobeying action $a_t = a^h - 1$. The payoff difference is bounded below by

$$-\int_{a^h-1}^{a^h} \Big[\psi F'(x) - C'(x)\Big] dx + H(-1; 1) - \bar{\lambda}_H K.$$

Because F' is decreasing and $\lim_{x\to\infty} C'(x) = \infty$, this lower bound approaches infinity. Continuity and strict monotonicity now imply that there are unique cutoffs

$$0 < \underline{a}(\psi, \bar{\lambda}_H K) < a^m(\psi) < \bar{a}(\psi, \bar{\lambda}_H K) < \infty \tag{20}$$

such that, under full belief, maximal disobedience gives at least as high a payoff as obedience

if and only if

$$a^h \leq \underline{a}(\psi, \bar{\lambda}_H K) \quad \text{or} \quad a^h \geq \bar{a}(\psi, \bar{\lambda}_H K). \tag{21}$$

For every $\mu_{t-1} < 1$, disobedience is strictly more attractive than under full belief. Indeed, for every $a_t \neq a^h$,

$$H(a_t - a^h; \mu_{t-1}) > H(a_t - a^h; 1)$$

and

$$-\mu_{t-1}\bar{\lambda}_H K > -\bar{\lambda}_H K.$$

Thus, under every commandment in Expression 21, the individual disobeys at every finite belief. This conclusion includes both cutoff commandments. After i sufferings, the posterior odds are

$$\frac{\mu(i)}{1-\mu(i)} = \frac{\mu_0}{1-\mu_0}\left(\frac{\bar{\lambda}_H}{\lambda_M}\right)^i. \tag{22}$$

Since $\bar{\lambda}_H > \lambda_M$, Equation 22 implies $\lim_{i\to\infty} \mu(i) = 1$. Every religion in Expression 21 is therefore stable.

Conversely, suppose

$$\underline{a}(\psi, \bar{\lambda}_H K) < a^h < \bar{a}(\psi, \bar{\lambda}_H K). \tag{23}$$

If $a^h = a^m(\psi)$, obedience maximizes both the secular payoff and the guilt payoff. The first suffering then gives

$$\mu(1) = \frac{\mu_0 \cdot 0}{\mu_0 \cdot 0 + (1-\mu_0)\lambda_M} = 0.$$

If $a^h \neq a^m(\psi)$, maximal disobedience gives a strictly lower payoff than obedience at $\mu_{t-1} = 1$. By continuity, obedience is strictly optimal whenever belief is sufficiently close to one. If the individual has not obeyed earlier, Equation 22 implies that finitely many sufferings while disobeying raise his belief above this level. He then chooses $a_t^* = a^h$. If no suffering occurs, his belief remains unchanged and he continues to obey. At the first subsequent suffering,

$$\mu_t = \frac{\mu_{t-1} \cdot 0}{\mu_{t-1} \cdot 0 + (1-\mu_{t-1})\lambda_M} = 0.$$

Thus every religion in Expression 23 is unstable. This proves the first part of the proposition.

It remains to establish the behavior of action under a stable religion. Equation 16 and the implicit function theorem give

$$\frac{da_t^*}{d\mu_{t-1}} = \frac{\frac{\partial^2 H}{\partial x \partial \mu}(a_t^* - a^h; \mu_{t-1})}{-\left[\psi F''(a_t^*) - C''(a_t^*) + \frac{\partial^2 H}{\partial x^2}(a_t^* - a^h; \mu_{t-1})\right]}. \tag{24}$$

The denominator is strictly positive. If $a^h < a^m(\psi)$, Equation 17 gives $a_t^* - a^h > 0$. The assumed cross derivative is then negative, so $da_t^*/d\mu_{t-1} < 0$. If $a^h > a^m(\psi)$, Equation 18 gives $a_t^* - a^h < 0$. The cross derivative is then positive, so $da_t^*/d\mu_{t-1} > 0$. Belief is unchanged when no suffering occurs and strictly increases when suffering occurs under disobedience. Hence

$$|a_{t+1}^* - a^h| \leq |a_t^* - a^h|$$

in every period. Equations 17 and 18 also show that $a_t^* \neq a^h$ at every finite belief. This proves the second part. □

Proposition 2. *1. A religion is unstable if and only if it sanctifies a "moderate" action as the commandment $a^h \geq 0$:*

$$a^h \in \Big(\underline{a}(\psi, \bar{\lambda}_H K), \bar{a}(\psi, \bar{\lambda}_H K)\Big), \tag{7}$$

where the cutoffs $\underline{a}(\psi, \bar{\lambda}_H K)$ and $\bar{a}(\psi, \bar{\lambda}_H K)$ are the same as in Proposition 1.

2. For an unstable religion, the long-run action is the secular optimal action. Specifically, there is a finite integer $I < \infty$, so that if the individual suffers more than I disasters,

$$a_t^* = a^m(\psi).$$

Proof. The first part follows from the proof of Proposition 1. A commandment is unstable if and only if

$$\underline{a}(\psi, \bar{\lambda}_H K) < a^h < \bar{a}(\psi, \bar{\lambda}_H K).$$

For every such commandment, the proof for Proposition 1 also shows that, after finitely many sufferings, the individual obeys, and that the first suffering during obedience produces a zero posterior belief.

By $H(x; 0) = 0$ for every x, once belief is zero the individual's payoff from action is

$$\psi F(a_t) - C(a_t) - \lambda_M K.$$

We thus have

$$a_t^* = \arg\max_{a_t > 0} \big[\psi F(a_t) - C(a_t)\big] = a^m(\psi).$$

Belief zero is absorbing under Bayes' rule. Hence there is a finite integer I such that, after the individual has experienced more than I disasters, every subsequent action equals $a^m(\psi)$. □

Proposition 3. *Under a higher material return ψ, a salvation religion a^h is more likely to be mystic and less likely to be activist.*

Proof. For either cutoff, denote a^* as the fully believing individual's unique optimal disobeying action evaluated at that cutoff. Applying the envelope theorem and the implicit function theorem to the indifference condition in Equation 19, we have

$$\frac{\partial \bar{a}(\psi, \bar{\lambda}_H K)}{\partial \psi} = \frac{F\big(\bar{a}(\psi, \bar{\lambda}_H K)\big) - F(a^*)}{\psi\big[F'(a^*) - F'\big(\bar{a}(\psi, \bar{\lambda}_H K)\big)\big] + C'\big(\bar{a}(\psi, \bar{\lambda}_H K)\big) - C'(a^*)} > 0. \tag{25}$$

The numerator is positive because

$$a^m(\psi) < a^* < \bar{a}(\psi, \bar{\lambda}_H K).$$

The denominator is also positive because $F'' < 0$ and $C'' > 0$. The same argument gives

$$\frac{\partial \underline{a}(\psi, \bar{\lambda}_H K)}{\partial \psi} = \frac{F\big(\underline{a}(\psi, \bar{\lambda}_H K)\big) - F(a^*)}{\psi\big[F'(a^*) - F'\big(\underline{a}(\psi, \bar{\lambda}_H K)\big)\big] + C'\big(\underline{a}(\psi, \bar{\lambda}_H K)\big) - C'(a^*)} > 0. \tag{26}$$

Here both the numerator and the denominator are negative because

$$\underline{a}(\psi, \bar{\lambda}_H K) < a^* < a^m(\psi).$$

Thus both cutoffs strictly increase with ψ.

The probability that a salvation religion is mystic is

$$\frac{L\big(\underline{a}(\psi, \bar{\lambda}_H K)\big)}{L\big(\underline{a}(\psi, \bar{\lambda}_H K)\big) + 1 - L\big(\bar{a}(\psi, \bar{\lambda}_H K)\big)}.$$

Its derivative has the same sign as

$$\begin{aligned} & l\big(\underline{a}(\psi, \bar{\lambda}_H K)\big) \frac{\partial \underline{a}(\psi, \bar{\lambda}_H K)}{\partial \psi} \Big[1 - L\big(\bar{a}(\psi, \bar{\lambda}_H K)\big)\Big] \\ & + L\big(\underline{a}(\psi, \bar{\lambda}_H K)\big) l\big(\bar{a}(\psi, \bar{\lambda}_H K)\big) \frac{\partial \bar{a}(\psi, \bar{\lambda}_H K)}{\partial \psi}. \end{aligned} \tag{27}$$

Because l is strictly positive on the support, both cutoffs lie in its interior, and both cutoffs strictly increase with ψ, the derivative 27 is strictly positive. The derivative of the probability that a salvation religion is activist is therefore strictly negative. □

Proposition 4. *The probability that $Y(a^*(\bar{\psi}, \tilde{a}^h)) < Y(a^*(\underline{\psi}, \hat{a}^h))$ increases with $\bar{\lambda}_H K$.*

In other words, when individuals care more about religious deliverance from suffering ($\bar{\lambda}_H K$ increases), it is more likely that the location with a higher economic return has a lower economic outcome, $Y(a^*(\bar{\psi}, \tilde{a}^h)) < Y(a^*(\underline{\psi}, \hat{a}^h))$, or a reversal of economic fortune.

Proof. We first show that the economic outcome strictly increases with the religious commandment. Differentiating Equation 9 with respect to a^h, we have

$$\frac{\partial a^*(\psi, a^h)}{\partial a^h} = \frac{H''\big(a^*(\psi, a^h) - a^h; 1\big)}{\psi F''\big(a^*(\psi, a^h)\big) - C''\big(a^*(\psi, a^h)\big) + H''\big(a^*(\psi, a^h) - a^h; 1\big)} \in (0, 1). \tag{28}$$

It follows from $Y' > 0$ that $Y(a^*(\psi, a^h))$ strictly increases with a^h.

We next examine how the religious motivation changes the two distributions in Equation 11. Differentiating the indifference condition that determines the upper cutoff gives

$$\frac{\partial \bar{a}(\underline{\psi}, \bar{\lambda}_H K)}{\partial(\bar{\lambda}_H K)} = \frac{1}{\underline{\psi}\big[F'(a^*) - F'\big(\bar{a}(\underline{\psi}, \bar{\lambda}_H K)\big)\big] + C'\big(\bar{a}(\underline{\psi}, \bar{\lambda}_H K)\big) - C'(a^*)} > 0, \tag{29}$$

where a^* is evaluated at the material return $\underline{\psi}$ and the commandment $\bar{a}(\underline{\psi}, \bar{\lambda}_H K)$. Similarly,

$$\frac{\partial \underline{a}(\bar{\psi}, \bar{\lambda}_H K)}{\partial(\bar{\lambda}_H K)} = \frac{1}{\bar{\psi}\big[F'(a^*) - F'\big(\underline{a}(\bar{\psi}, \bar{\lambda}_H K)\big)\big] + C'\big(\underline{a}(\bar{\psi}, \bar{\lambda}_H K)\big) - C'(a^*)} < 0, \tag{30}$$

where a^* is now evaluated at $\bar{\psi}$ and $\underline{a}(\bar{\psi}, \bar{\lambda}_H K)$.

For any x in the interior of the Western support, the cumulative distribution of the religious commandment is

$$\Pr(\hat{a}^h \leq x) = \frac{L(x) - L\big(\bar{a}(\underline{\psi}, \bar{\lambda}_H K)\big)}{1 - L\big(\bar{a}(\underline{\psi}, \bar{\lambda}_H K)\big)}.$$

Its derivative with respect to $\bar{\lambda}_H K$ is

$$-\frac{[1 - L(x)]l\big(\bar{a}(\underline{\psi}, \bar{\lambda}_H K)\big)}{\big[1 - L\big(\bar{a}(\underline{\psi}, \bar{\lambda}_H K)\big)\big]^2} \frac{\partial \bar{a}(\underline{\psi}, \bar{\lambda}_H K)}{\partial(\bar{\lambda}_H K)} < 0.$$

Thus a stronger religious motivation shifts the distribution of $\hat{a}^h$ toward higher commandments.

For any x in the interior of the Eastern support, the corresponding conditional cumulative distribution is

$$\Pr(\tilde{a}^h \leq x) = \frac{L(x)}{L\big(\underline{a}(\bar{\psi}, \bar{\lambda}_H K)\big)}.$$

Its derivative is

$$-\frac{L(x)l\big(\underline{a}(\bar{\psi}, \bar{\lambda}_H K)\big)}{L\big(\underline{a}(\bar{\psi}, \bar{\lambda}_H K)\big)^2} \frac{\partial \underline{a}(\bar{\psi}, \bar{\lambda}_H K)}{\partial(\bar{\lambda}_H K)} > 0.$$

Thus a stronger religious motivation shifts the distribution of $\tilde{a}^h$ toward lower command-

ments.

By Equation 28, the Western economic outcome shifts upward and the Eastern economic outcome shifts downward. The two commandments are independent. First-order stochastic dominance therefore implies that

$$\Pr\Big(Y(a^*(\bar{\psi},\tilde{a}^h)) < Y(a^*(\underline{\psi},\hat{a}^h))\Big)$$

weakly increases with $\bar{\lambda}_H K$. The increase is strict whenever

$$Y\Big(a^*\big(\underline{\psi},\bar{a}(\underline{\psi},\bar{\lambda}_H K)\big)\Big) < Y\Big(a^*\big(\bar{\psi},\underline{a}(\bar{\psi},\bar{\lambda}_H K)\big)\Big).$$

If the reverse weak inequality holds, the reversal of economic fortunes already occurs with probability one, so the probability cannot increase further. □

Proposition 8. *Denote K_W as the cost of suffering in the Western location, and K_E as the cost of suffering in the Eastern location. Suppose that the religion in the Western location is activist, while the Eastern religion can be mystic or a mix between mysticism or activism. The probability that $Y(a^*(\bar{\psi},\tilde{a}^h)) < Y(a^*(\underline{\psi},\hat{a}^h))$ increases with K_W.*

Proof. By Equation 29 and $\bar{\lambda}_H > 0$,

$$\frac{\partial \bar{a}(\underline{\psi},\bar{\lambda}_H K_W)}{\partial K_W} > 0.$$

For any $x \in \big(\bar{a}(\underline{\psi},\bar{\lambda}_H K_W), \bar{A}^h\big)$, the distribution of the Western commandment satisfies

$$\Pr(\hat{a}^h \le x) = \frac{L(x) - L\big(\bar{a}(\underline{\psi},\bar{\lambda}_H K_W)\big)}{1 - L\big(\bar{a}(\underline{\psi},\bar{\lambda}_H K_W)\big)}.$$

Differentiating this expression gives

$$\frac{\partial \Pr(\hat{a}^h \le x)}{\partial K_W} = -\frac{[1-L(x)]l\big(\bar{a}(\underline{\psi},\bar{\lambda}_H K_W)\big)}{\big[1 - L\big(\bar{a}(\underline{\psi},\bar{\lambda}_H K_W)\big)\big]^2}\frac{\partial \bar{a}(\underline{\psi},\bar{\lambda}_H K_W)}{\partial K_W} < 0.$$

The distribution function is zero below its lower support. Therefore, a larger K_W shifts the Western commandment upward in the sense of first-order stochastic dominance. Equation 28 and $Y' > 0$ imply that $Y(a^*(\underline{\psi},\hat{a}^h))$ strictly increases with $\hat{a}^h$. Thus, the Western economic outcome also shifts upward in the sense of first-order stochastic dominance.

The distribution of the Eastern economic outcome does not change with K_W. Conditional on any realization of the Eastern commandment, independence implies that the probability

that the Western economic outcome is larger weakly increases with K_W. The result is proven by averaging over the unchanged distribution of the Eastern commandment.

□

Proposition 5. *1. For $a^h \in [a^m, \infty)$, there exists a unique cutoff $\bar{a}(\bar{\lambda}_H K) \in (a^m, \infty)$, such that a religion is stable if and only if $a^h \geq \bar{a}(\bar{\lambda}_H K)$.*

Suppose that w_s is large enough and η is small enough. For $a^h \in [0, a^m]$, there exists a unique cutoff $\underline{a}(\bar{\lambda}_H K) \in (0, a^m)$, such that a religion is stable if and only if $a^h \leq \underline{a}(\bar{\lambda}_H K)$.

2. Consider $a^h > \bar{a}(\bar{\lambda}_H K)$. Whenever the king persecutes the religio-cultural elite in a period t, the elite resists the king even more fiercely in the next period at $a^*_{t+1} > a^*_t$, forcing the king to choose a weaker domination in the next period at $D^*_{t+1} < D^*_t$.

Proof. Because $D' < 0$, $Q' < 0$, $D'' \geq 0$, $Q'' \geq 0$, and $C''_e > 0$,

$$\frac{d^2}{da^2} w_s[-D(a)Q(a) - \eta C_e(a)] = w_s[-D''(a)Q(a) - 2D'(a)Q'(a) - D(a)Q''(a) - \eta C''_e(a)] < 0. \tag{31}$$

Thus, the secular payoff has a unique maximizer a^m. Conditional on disobedience, the elite also has a unique optimal action. If $a^h > a^m$, then $a^m < a^*_t < a^h$; if $a^h < a^m$, then $a^h < a^*_t < a^m$. These inequalities follow because the marginal secular payoff is zero at a^m, the marginal guilt is zero at a^h, and guilt always pulls the action toward a^h.

We first consider $a^h \in [a^m, \infty)$. Under full belief, the maximal payoff from disobedience relative to obedience is

$$\begin{aligned} &\max_{a \neq a^h} \{w_s[-D(a)Q(a) - \eta C_e(a)] + H(a - a^h; 1) - \bar{\lambda}_H K\} \\ &\quad - w_s[-D(a^h)Q(a^h) - \eta C_e(a^h)]. \end{aligned} \tag{32}$$

For $a^h > a^m$, the envelope theorem and the first-order condition for a^*_t imply that the derivative of Expression 32 is

$$\begin{aligned} &w_s[-D'(a^*_t)Q(a^*_t) - D(a^*_t)Q'(a^*_t) - \eta C'_e(a^*_t)] \\ &\quad - w_s[-D'(a^h)Q(a^h) - D(a^h)Q'(a^h) - \eta C'_e(a^h)] > 0, \end{aligned}$$

where the inequality follows from $a^*_t < a^h$ and strict concavity. As $a^h \downarrow a^m$, Expression 32 converges to $-\bar{\lambda}_H K < 0$. For $a^h > 1$, choosing $a^h - 1$ instead of obeying gives the payoff

difference

$$w_s[-D(a^h-1)Q(a^h-1)-\eta C_e(a^h-1)+D(a^h)Q(a^h)+\eta C_e(a^h)] \\ +H(-1;1)-\bar{\lambda}_H K.$$

This expression converges to infinity because $D(a)Q(a)$ is bounded and $\lim_{a\to\infty} C'_e(a)=\infty$. Continuity and strict monotonicity therefore give a unique $\bar{a}(\bar{\lambda}_H K)\in(a^m,\infty)$.

For $a^h \geq \bar{a}(\bar{\lambda}_H K)$, disobedience is weakly optimal under full belief and strictly optimal at every finite belief, since a lower belief weakens guilt and reduces the expected punishment from disobedience. After i persecutions while disobeying,

$$\frac{\mu(i)}{1-\mu(i)}=\frac{\mu_0}{1-\mu_0}\left(\frac{\bar{\lambda}_H}{\lambda_M}\right)^i\to\infty.$$

Thus, the religion is stable. For $a^h\in(a^m,\bar{a}(\bar{\lambda}_H K))$, obedience is strictly optimal when belief is sufficiently close to one. Finitely many persecutions while disobeying induce obedience, and the first persecution during obedience sets belief equal to zero. The same conclusion holds at $a^h=a^m$. Hence, on $[a^m,\infty)$, stability is equivalent to $a^h\geq\bar{a}(\bar{\lambda}_H K)$.

We next consider $a^h\in[0,a^m]$. The restrictions on w_s and η are used only to ensure

$$\max_{a>0}\{w_s[-D(a)Q(a)-\eta C_e(a)]+H(a;1)-\bar{\lambda}_H K\}>w_s[-D(0)Q(0)-\eta C_e(0)]. \tag{33}$$

Indeed, fix any $a>0$. Since $D(a)Q(a)<D(0)Q(0)$, choosing η sufficiently small makes the secular payoff at a strictly higher than at zero. Choosing w_s sufficiently large then makes this secular gain larger than $\bar{\lambda}_H K-H(a;1)$, which proves Condition 33.

Condition 33 makes Expression 32 strictly positive at $a^h=0$. For $0<a^h<a^m$, its envelope derivative is strictly negative because $a^h<a^*_t$. As $a^h\uparrow a^m$, the expression converges to $-\bar{\lambda}_H K<0$. Hence, there is a unique $\underline{a}(\bar{\lambda}_H K)\in(0,a^m)$. The belief argument above then shows that the religion is stable if and only if $a^h\leq\underline{a}(\bar{\lambda}_H K)$.

Finally, suppose $a^h>\bar{a}(\bar{\lambda}_H K)$. Whenever the elite is persecuted, his belief strictly increases. His action satisfies

$$w_s[-D'(a^*_t)Q(a^*_t)-D(a^*_t)Q'(a^*_t)-\eta C'_e(a^*_t)]+H'(a^*_t-a^h;\mu_{t-1})=0.$$

The denominator in the implicit-function derivative is the negative of the second derivative

of the objective and is strictly positive. Its numerator is

$$\frac{\partial^2 H}{\partial x \partial \mu}(a_t^* - a^h; \mu_{t-1}) > 0$$

because $a_t^* < a^h$. Thus $da_t^*/d\mu_{t-1} > 0$. Every persecution therefore implies $a_{t+1}^* > a_t^*$ and, because $D' < 0$, $D_{t+1}^* < D_t^*$. □

Proposition 6. *Suppose that the Mandarin's initial beliefs satisfy*

$$\mu_0^c \geq \mu_H, \qquad 0 < \mu_0^d \leq \mu_0^c \mu_H, \qquad 1 - \mu_0^c - \mu_0^d > 0.$$

Then there exist finite integers $I \geq 0$ and $J > I + 1$ such that:

1. *The Mandarin first obeys the Confucian commandment after I realized persecutions. Each of the first I persecutions implies $a_{t+1}^* > a_t^*$ and $D_{t+1}^* < D_t^*$.*

2. *At the $(I+1)$-st persecution, the Mandarin is obeying the Confucian commandment. Immediately afterward, his posterior belief in Confucianism is zero.*

3. *After more than J persecutions, each new persecution implies $a_{t+1}^* < a_t^*$ and $D_{t+1}^* > D_t^*$.*

Proof. We first consider the phase before the Mandarin obeys the Confucian commandment. We show by induction that he disobeys both commandments during this phase. Suppose that both commandments were disobeyed before each of the first i persecutions. Denote $\mu^c(i)$ and $\mu^d(i)$ as beliefs after i realized persecutions. Bayes' rule gives

$$\frac{\mu^d(i)}{\mu^c(i)} = \frac{\mu_0^d}{\mu_0^c} \left(\frac{\bar{\lambda}_D}{\bar{\lambda}_C} \right)^i \leq \mu_H$$

and

$$\frac{1 - \mu^c(i) - \mu^d(i)}{\mu^c(i)} = \frac{1 - \mu_0^c - \mu_0^d}{\mu_0^c} \left(\frac{\lambda_M}{\bar{\lambda}_C} \right)^i.$$

Thus $\mu^c(i)$ strictly increases and converges to one. Moreover,

$$\mu^d(i) < \frac{\mu^d(i)}{\mu^c(i)} \leq \mu_H,$$

so Taoist guilt is zero throughout this phase.

The payoff from choosing a^c minus the payoff from choosing a^d is

$$w_s\left\{[-D(a^c)Q(a^c)-\eta C_e(a^c)]-[-D(a^d)Q(a^d)-\eta C_e(a^d)]\right\}$$
$$-H(a^d-a^c;\mu^c(i))+[\mu^c(i)\bar{\lambda}_C-\mu^d(i)\bar{\lambda}_D]K>0.$$

The first term is positive by Assumption 1; the second is nonnegative; and the last is positive because $\mu^d(i)<\mu^c(i)$ and $\bar{\lambda}_D<\bar{\lambda}_C$. Thus a^d is not optimal. If a^c is also not chosen, both commandments are disobeyed before the next persecution. Since the same argument applies at $i=0$, this completes the induction.

Whenever $\mu^c(i)>\mu_H$ and the Mandarin has not yet chosen a^c, strict concavity gives a unique action satisfying

$$a^m<a_t^*<a^c$$

and

$$w_s[-D'(a_t^*)Q(a_t^*)-D(a_t^*)Q'(a_t^*)-\eta C_e'(a_t^*)]+H'(a_t^*-a^c;\mu^c(i))=0.$$

The denominator in the implicit-function derivative is the negative of the second derivative of the Mandarin's objective and is strictly positive by Equation 31. Its numerator is

$$\frac{\partial^2 H}{\partial x\partial\mu}(a_t^*-a^c;\mu^c(i))>0.$$

Hence the Mandarin's resistance strictly increases with his Confucian belief. Every persecution in this phase therefore increases his resistance and decreases the emperor's optimal domination. If $\mu_0^c=\mu_H$ and a^c is not chosen immediately, both guilt terms are initially zero and the Mandarin chooses a^m. The first persecution raises Confucian belief strictly above μ_H, so his next action is strictly above a^m.

Because

$$a^m<a^c<\bar{a}(\bar{\lambda}_C K),$$

obedience to a^c is strictly optimal when Confucian belief is sufficiently close to one and Taoist belief is sufficiently close to zero. The Bayesian expressions above show that these beliefs are reached after finitely many persecutions. Let I be the number of persecutions before a^c is first chosen. At the persecution that induces this first obedience, the action rises from a value below a^c to a^c. This proves Part 1, including the case $I=0$ in which a^c is chosen immediately.

At the next persecution, the Mandarin is obeying a^c. Bayes' rule therefore gives

$$\mu^c=\frac{\mu^c(I)\cdot 0}{\mu^d(I)\bar{\lambda}_D+[1-\mu^c(I)-\mu^d(I)]\lambda_M}=0.$$

The denominator is strictly positive. The Taoist posterior immediately after this persecution is

$$\frac{\mu^d(I)\bar{\lambda}_D}{\mu^d(I)\bar{\lambda}_D + [1-\mu^c(I)-\mu^d(I)]\lambda_M},$$

which is strictly between zero and one. This proves Part 2.

After Confucian belief becomes zero, Confucian guilt is zero. The lower cutoff is strictly decreasing in the expected punishment from disobedience. This follows directly by implicit differentiation of the cutoff indifference condition: the maximal payoff from disobedience relative to obedience is strictly decreasing in the commandment below a^m, while a larger punishment reduces this payoff difference one for one. Hence

$$\underline{a}(\bar{\lambda}_C K) < \underline{a}(\bar{\lambda}_D K).$$

Assumption 1 implies

$$a^d < \underline{a}(\bar{\lambda}_D K).$$

The Taoist commandment is therefore a stable mystic commandment, and the Mandarin never chooses it exactly.

After every subsequent persecution, the posterior odds of Taoism relative to the naturalist worldview are multiplied by $\bar{\lambda}_D/\lambda_M > 1$. Both worldviews have strictly positive belief immediately after Confucian belief becomes zero and after every finite number of later persecutions. Taoist belief therefore strictly increases and converges to one. Since $\mu_H < 1$, there is a finite $J > I+1$ such that Taoist belief is strictly greater than μ_H after J persecutions.

After that point, the Mandarin's unique action satisfies

$$a^d < a_t^* < a^m$$

and

$$w_s[-D'(a_t^*)Q(a_t^*) - D(a_t^*)Q'(a_t^*) - \eta C_e'(a_t^*)] + H'(a_t^* - a^d; \mu^d) = 0.$$

The denominator in the implicit-function derivative is again strictly positive. Its numerator is

$$\frac{\partial^2 H}{\partial x \partial \mu}(a_t^* - a^d; \mu^d) < 0.$$

Thus resistance strictly decreases with Taoist belief. Every persecution after the first J persecutions strictly decreases the Mandarin's resistance. Since $D' < 0$, the emperor's optimal domination strictly increases. This proves Part 3. □

Proposition 7. *There is a unique symmetric equilibrium in the commoners' efforts in the*

status contest, denoted as e_t^*.

1. *Suppose that* a^h *is a world-adjusting religion:* $a^h \in \left(\underline{a}(\psi, \bar{\lambda}_H K), \bar{a}(\psi, \bar{\lambda}_H K)\right)$ *as defined in Proposition 2.*

   (a) *Every time the representative commoner suffers, the more believing commoner exerts a higher effort in the contest.*

   (b) *The long-run effort* $\lim_{t\to\infty} e_t^*$ *is strictly positive after a sufficiently large number of sufferings. The long-run effort strictly increases with* a^h *if* $a^h < a^m(\psi)$ *and strictly decreases with* a^h *if* $a^h > a^m(\psi)$.

2. *Suppose that* a^h *is a salvation religion:* $a^h \in [0,\infty)\backslash\left(\underline{a}(\psi, \bar{\lambda}_H K), \bar{a}(\psi, \bar{\lambda}_H K)\right)$ *as in Proposition 1. Then commoners always exert zero efforts in the contest:* $e_t^* = 0$.

Proof. We use the standard boundary conditions

$$\Gamma'(0) = 0, \qquad \frac{\partial \Pi(0;0)}{\partial e_t^i} > 0. \tag{34}$$

As long as an individual remains a commoner, the religious worldview assigns probability $\bar{\lambda}_H$ to suffering independently of his action. Hence, after suffering,

$$\mu_t = \frac{\mu_{t-1}\bar{\lambda}_H}{\mu_{t-1}\bar{\lambda}_H + (1-\mu_{t-1})\lambda_M} > \mu_{t-1}.$$

If a commoner fails in the contest, his action is the unique solution to

$$a_t^* \in \arg\max_{a\geq 0} \left\{\psi F(a) - C(a) + H(a - a^h; \mu_{t-1})\right\}.$$

If an elite does not choose a^h, elite status changes neither his action nor his payoff. If the elite chooses a^h, the value of elite status is

$$\begin{aligned} &\psi F(a^h) - C(a^h) - \left[\psi F(a_t^*) - C(a_t^*) + H(a_t^* - a^h; \mu_{t-1})\right] \\ &\quad + \mu_{t-1}\bar{\lambda}_H K. \end{aligned} \tag{35}$$

Thus, up to terms independent of effort, commoner i solves

$$\begin{aligned} \max_{e_t^i \geq 0} \quad & -\Gamma(e_t^i) + \Pi(e_t^i; e_t^{-i}) \\ & \times \max\left\{0, \psi F(a^h) - C(a^h) - \left[\psi F(a_t^*) - C(a_t^*) + H(a_t^* - a^h; \mu_{t-1})\right] + \mu_{t-1}\bar{\lambda}_H K\right\}. \end{aligned} \tag{36}$$

The objective is strictly concave in e_t^i. When the expression inside the maximum is zero, the unique optimal effort is zero. When it is positive, the symmetric equilibrium satisfies

$$\begin{aligned}\Gamma'(e_t^*) =& \frac{\partial \Pi(e_t^*; e_t^*)}{\partial e_t^i} \Big\{ \psi F(a^h) - C(a^h) \\ & - \Big[\psi F(a_t^*) - C(a_t^*) + H(a_t^* - a^h; \mu_{t-1}) \Big] + \mu_{t-1} \bar{\lambda}_H K \Big\}. \end{aligned} \tag{37}$$

The derivative of the left side minus the right side with respect to symmetric effort is

$$\begin{aligned}&\Gamma''(e_t^*) - \left[\frac{\partial^2 \Pi(e_t^*; e_t^*)}{\partial (e_t^i)^2} + \frac{\partial^2 \Pi(e_t^*; e_t^*)}{\partial e_t^i \partial e_t^{-i}} \right] \\ &\quad \times \Big\{ \psi F(a^h) - C(a^h) - \Big[\psi F(a_t^*) - C(a_t^*) + H(a_t^* - a^h; \mu_{t-1}) \Big] + \mu_{t-1} \bar{\lambda}_H K \Big\} > 0. \end{aligned}$$

The symmetric equilibrium therefore exists and is unique.

Before an elite chooses a^h, elite status has zero value and equilibrium effort is zero. Once an elite chooses a^h, the envelope theorem shows that Expression 35 has derivative

$$\bar{\lambda}_H K - \frac{\partial H(a_t^* - a^h; \mu_{t-1})}{\partial \mu_{t-1}} > 0$$

with respect to belief. Hence equilibrium effort weakly increases after every suffering and strictly increases after every suffering once elite status has positive value.

Suppose now that

$$\underline{a}(\psi, \bar{\lambda}_H K) < a^h < \bar{a}(\psi, \bar{\lambda}_H K).$$

At full belief, obedience is strictly better than disobedience for an elite. By continuity, the value of elite status is strictly positive after a sufficiently large number of sufferings. Since $\Gamma'(0) = 0$ and

$$\frac{\partial \Pi(0; 0)}{\partial e_t^i} > 0,$$

the unique symmetric equilibrium effort is then strictly positive.

It remains to establish the comparative statics with respect to a^h. At full belief, the derivative of Expression 35 is

$$\Big[\psi F'(a^h) - C'(a^h) \Big] - \Big[\psi F'(a_t^*) - C'(a_t^*) \Big].$$

If $a^h < a^m(\psi)$, then $a^h < a_t^* < a^m(\psi)$, so this derivative is strictly positive. If $a^h > a^m(\psi)$, then $a^m(\psi) < a_t^* < a^h$, so it is strictly negative. Because equilibrium effort is strictly increasing in the value of elite status, long-run effort strictly increases with a^h below $a^m(\psi)$

and strictly decreases with a^h above $a^m(\psi)$.

Finally, suppose that

$$a^h \leq \underline{a}(\psi, \bar{\lambda}_H K) \quad \text{or} \quad a^h \geq \bar{a}(\psi, \bar{\lambda}_H K).$$

At every finite belief, an elite strictly prefers disobedience. Elite status then changes neither the action nor the payoff, so its value is zero and the unique equilibrium effort is

$$e_t^* = 0.$$

This conclusion includes both cutoff commandments. □

B Bayesian religiosity and the legibility problem

In this appendix, individuals update their beliefs under both sufferings and good fortunes, hence fully Bayesian; a just individual may also suffer even under the religious worldview. Through this more general framework, we show that results of our baseline model are not driven by asymmetric updating. Our results are robust as long as Bayesian individuals, *in the long run, suffer more than they enjoy good fortunes.*

This appendix also uncovers new insights. While the sufferers are always attracted to strict religions, world adjustment does not have reliable bases of followers. Consequently, it is far more difficult to infer religiosity under world adjustment. This problem of legibility bridges our analysis with prominent economic theories of churches on the supply side (e.g., Iannaccone (1992)).

B.1 The setup revisited

The worldviews and the payoff function. Under the religious worldview, the world is just. The probability of suffering is

$$\lambda_H = \begin{cases} \underline{\lambda} & \text{if } a_t = a^h \\ \bar{\lambda} & \text{if } a_t \neq a^h \end{cases} \tag{B.1}$$

In Equation B.1, $\bar{\lambda}$ is the probability of suffering for an unjust individual and $\underline{\lambda}$ is the probability of suffering for a just/moral individual. Assume that $0 < \underline{\lambda} < \lambda_M$, where λ_M is the probability of suffering under the mechanical worldview. A just individual thus suffers less in a just world than in a morally indifferent world. We also keep $\lambda_M < \bar{\lambda} < 1$, as in the baseline model: an unjust individual suffers more in a just world than in a morally indifferent world.

Thus, to the payoff function in Equation 3, we need to add an extra term $-\mu_{t-1}\mathbf{1}\{a_t = a^h\}\underline{\lambda}K$, which is the expected suffering of a just individual under the religious worldview. The individual now uses Bayes' rule to update belief under *both* suffering and good fortunes.

The process of fortunes. This section needs to explicitly work with the stochastic process of suffering. Denote

$$\theta_t \in \{0, 1\} \text{ and } \theta := (\theta_1, \theta_2, ..., \theta_t, ...),$$

where $\theta_t = 1$ if a bad fortune (i.e., suffering) strikes in period t, and $\theta_t = 0$ if good fortune happens (i.e., no suffering). The stochastic process θ is defined on the space $(\Omega, \mathcal{H}, \mathbb{P})$; for an $\omega \in \Omega$, $\theta(\omega)$ is a realized path of fortunes. Denote $\{\mathcal{F}_t\}_{t=0}^{\infty}$ as the canonical filtration, where

each $\mathcal{F}_t$ is the σ-algebra generated by outcomes $\{\theta_s\}_{s=1}^t$ and actions $\{a_s\}_{s=1}^t$. Given our setup, action a_t is $\mathcal{F}_{t-1}$-measurable. The objective probability of bad fortune is $\mathbb{P}(\theta_t = 1|\mathcal{F}_{t-1})$, denoted as p_t. Assume that

$$p_t = p(a_t; \mathcal{F}_{t-1}) \in (0,1) \text{ for all } t,$$

which also implies that p_t is $\mathcal{F}_{t-1}$-measurable. Also denote

$$i_t(\theta) := \sum_{s=1}^{t} \theta_s \text{ and } j_t(\theta) := t - i_t(\theta),$$

where i_t is the number of bad fortunes and j_t is the number of good fortunes until period t. The ratio j_t/i_t measures how fortunate an individual is, which may also proxy social classes.

Definition B.1. *Fix an $\alpha > 0$. Define the following set of processes:*

$$\Theta(\alpha) := \Big\{\theta(\omega) : \textit{as } t \to \infty, i_t \to \infty \textit{ and } \frac{j_t}{i_t} \to \alpha\Big\}.$$

An individual with α belongs to the social base of a religion a^h if and only if for any $\mu_0 \in (0,1)$ and under any $\theta(\omega) \in \Theta(\alpha)$, $\mu_t \to 1$.

Suppose that an individual's process of fortunes belongs to the set $\Theta(\alpha)$. Then asymptotically, the ratio of fortunes j_t/i_t converges to α. Under a large α, the individual largely experiences good fortunes in the long run. Individuals with a specific α belong to the "base" of a religion a^h if the belief μ_t converges to 1 for all processes in $\Theta(\alpha)$, starting from any initial belief $\mu_0 \in (0,1)$.

As our main objective, we want to characterize the long-run belief for different α. The following lemma narrows down the possibilities.

Lemma B.1. *For any process $\theta(\omega)$ and any $\hat{\mu} \in (0,1)$, as $t \to \infty$, $\mu_t \not\to \hat{\mu}$. In other words, the belief μ_t cannot converge to an interior value in $(0,1)$.*

All proofs for Appendix B are in Appendix B.5. Because belief cannot converge to any interior belief $\hat{\mu} \in (0,1)$, there are only three possibilities: the belief converges to 0 or 1, or $\lim_{t\to\infty} \mu_t$ does not exist.

B.2 The social base and high legibility of strict religions

To better characterize religions, consider the difference between the maximal payoff from a balance between material and guilt motives and the payoff from the just action $a_t = a^h$:

$$\Delta\big(\mu_{t-1}; a^h, \psi, (\bar{\lambda} - \underline{\lambda})K\big) := \max_{a \geq 0}[\psi F(a) - C(a) + H(a - a^h; \mu_{t-1})]$$

$$-\Big\{[\psi F(a^h) - C(a^h)] + [\mu_{t-1}(\bar{\lambda} - \underline{\lambda})K]\Big\}. \tag{B.2}$$

The individual then chooses $a_t = a^h$ if and only if $\Delta\big(\mu_{t-1}; a^h, \psi, (\bar{\lambda} - \underline{\lambda})K\big) < 0$. We first focus on the following set of religions:

$$\mathcal{D} := \Big\{a \geq 0 : \Delta\big(1; a, \psi, (\bar{\lambda} - \underline{\lambda})K\big) \geq 0\Big\}. \tag{B.3}$$

For any religion in the set $\mathcal{D}$, the individual disobeys the religious commandment under any belief. As in the baseline model, the set $\mathcal{D}$ satisfies the following cutoff property:

Lemma B.2. *A necessary and sufficient condition for $a^h \in \mathcal{D}$ is: $a^h \leq \underline{a}\big(\psi, (\bar{\lambda} - \underline{\lambda})K\big)$ or $a^h \geq \bar{a}\big(\psi, (\bar{\lambda} - \underline{\lambda})K\big)$, under the cutoffs $\underline{a}\big(\psi, (\bar{\lambda} - \underline{\lambda})K\big) < a^m(\psi)$ and $\bar{a}[\psi, (\bar{\lambda} - \underline{\lambda})K] > a^m(\psi)$.*

The lemma is immediately established by replacing $\bar{\lambda}_H K$ in the baseline model with $(\bar{\lambda} - \underline{\lambda})K$. The set $\mathcal{D}$ is the set of strict religions that are beyond human capability. The following proposition identifies their long-run belief, which is determined by the following cutoff for α:

$$\alpha_0 := \frac{\log(\frac{\bar{\lambda}}{\lambda_M})}{\log(\frac{1-\lambda_M}{1-\bar{\lambda}})}. \tag{B.4}$$

Proposition B.1. *Suppose that $a^h \in \mathcal{D}$. As $t \to \infty$, the following statements are true.*

1. *If $\alpha < \alpha_0$, then under any $\mu_0 \in (0,1)$ and any process $\theta(\omega) \in \Theta(\alpha)$, $\mu_t \to 1$.*
2. *If $\alpha > \alpha_0$, then under any $\mu_0 \in (0,1)$ and any process $\theta(\omega) \in \Theta(\alpha)$, $\mu_t \to 0$.*

In Proposition B.1, the long-run belief is determined by α, the asymptotic ratio of good fortunes to bad fortunes. With $\alpha < \alpha_0$, these unfortunate individuals attain a strong conviction in the religion a^h, regardless of their initial belief μ_0 and the specific sequence of fortunes. Under $\alpha > \alpha_0$, these fortunate individuals are eventually fully disenchanted.

The mechanism is familiar. Under $a^h \in \mathcal{D}$, the individual never obeys the commandment a^h. If bad fortunes are sufficiently more frequent than good fortunes in the long run, the religion fully convinces the individual: as a "sinner," he is largely suffering in his life, which

confirms the religious worldview. On the flip side, more frequent good fortunes will induce full disenchantment because the individual is largely enjoying his life as a sinner. Thus, for the baseline model, its main insight on strict religions is not driven by asymmetric updating, but by the *eventual* dominance of misfortunes in one's life experiences.

The high legibility of religiosity under strict religions As an important consequence of Proposition B.1, the long-run religiosity of an individual requires only minimal information to estimate, therefore highly legible to its potential "church." Such a church only needs to know whether the ratio α is below α_0. This high legibility of religiosity is supported by the unattainable commandment, which generates a process of belief updating that is *only determined by the total number of bad and good fortunes*. The church thus does not need to know early life experiences, such as initial belief μ_0 or the precise sequence of fortunes. These early life experiences, while difficult to observe, are informative on long-run belief in world-adjusting religions, making inference of religiosity substantially more challenging (Section B.3). [12]

B.3 The less fortunate and world adjustment

To analyze world-adjusting religions, consider the following set of a^h:

$$\mathcal{A} := \Big\{ a \geq 0 : \Delta\big(1; a, \psi, (\bar{\lambda} - \underline{\lambda})K\big) < 0 \text{ and } a \neq a^m(\psi) \Big\}.$$

The case of $a^h = a^m(\psi)$ is knife-edge and special, so we analyze it separately in Appendix C.5. As a simple but key feature of any $a^h \in \mathcal{A}$, there is a unique cutoff $\underline{\mu}$, such that the individual obeys for $\mu_{t-1} > \underline{\mu}$ and still disobeys for $\mu_{t-1} \leq \underline{\mu}$.[13]Because these religions have both the obey and disobey regimes, it is necessary to consider two cutoffs for α, the asymptotic likelihood ratio of fortunes. The first one is still α_0, which controls the long-run direction of belief updating *in the disobey regime* (Section B.2). The second one is

$$\alpha_1 := \frac{\log(\frac{\lambda_M}{\underline{\lambda}})}{\log(\frac{1-\underline{\lambda}}{1-\lambda_M})}. \tag{B.5}$$

The second cutoff α_1 controls the long-run direction of belief updating *in the obey regime*. It turns out that $\alpha_0 < \alpha_1$ (see proof in Lemma B.4). This suggests that, under world

[12]Action also informs long-run belief; but in the context of religion, action is also very difficult to observe for a church because of the strong incentive for followers to hide immoral action (Norenzayan (2013)). The enormous difficulty to observe immoral action is actually the central explanation from the Big Gods literature on why religions need to postulate supernatural monitoring in the first place (Norenzayan (2013)).

[13]This intuitive result is proven by Lemma B.3 in Appendix B.5.

adjustment, the population should be demarcated into three groups. Those with $\alpha < \alpha_0$ are the most unfortunate, the same as in strict religions. Those with $\alpha > \alpha_1$ are now the (most) "fortunate," and $\alpha_0 < \alpha < \alpha_1$ are the "marginally fortunate." We consider the three cases one by one.

Proposition B.2. *If $\alpha < \alpha_0$, for any $\mu_0 \in (0,1)$ and $\theta(\omega) \in \Theta(\alpha)$, $\lim_{t\to\infty} \mu_t$ does not exist.*

Since the limit does not exist and μ_t is bounded, the path of μ_t must permanently oscillate. The oscillation is driven by the existence of both the obey regime and the disobey regime. For an obeying individual who suffers a lot, his belief in religion declines, eventually dropping below the threshold

$$\underline{\mu} \in (0,1).$$

Under $\mu_{t-1} \leq \underline{\mu}$, the individual disobeys, but then suffering will propel him to believe more in religion, which will rise above the threshold $\underline{\mu} \in (0,1)$ and will drive the individual to obey again, restarting the cycle.

Proposition B.2 shows that it is impossible to know the precise long-run belief of the unfortunate. The unfortunate individuals cannot be the social bases of world adjustment. The next proposition looks at the "marginally fortunate."

Proposition B.3. *1. Suppose that $\alpha < \alpha_1$. For any $\mu_0 \in (0,1)$ and any $\theta(\omega) \in \Theta(\alpha)$, either $\lim_{t\to\infty} \mu_t$ does not exist or $\lim_{t\to\infty} \mu_t = 0$.*

Because of Lemma B.1, we only need to rule out $\lim_{t\to\infty} \mu_t = 1$. Indeed, under $\alpha < \alpha_1$, good fortunes arrive too rarely so that an obeying individual will reduce his belief to zero, contradicting $\lim_{t\to\infty} \mu_t = 1$. The marginally fortunate ($\alpha_0 < \alpha < \alpha_1$) is not a social base of world adjustment either.

B.4 The fortunate under world adjustment

For any initial belief $\mu_0 \in (0,1)$ is also straightforward to construct a process of fortune $\theta(\omega)$, so that $\mu_t \to 0$ even under $\alpha > \alpha_1$. This shows the existence of non-believers among the fortunate, but not how likely we are to encounter non-believers. To do so, we impose more structures on the stochastic processes of fortunes, which allow us bound the probability of belief convergence.

Consider fortunate individuals who experience more good fortunes not only asymptotically, but likely so in every period. For such an individual, assume that there exists a (fixed but arbitrarily small) $\varepsilon > 0$, so that

$$\frac{1-p_t}{p_t} \geq \alpha_1 + \varepsilon \text{ for all } t. \tag{B.6}$$

The objective ratio of fortunes is therefore always larger than $\alpha_1 + \varepsilon$, not only asymptotically.[14] We can establish the following proposition, which uncovers a full belief bifurcation for the fortunate under world adjustment.

Proposition B.4. *1. Denote $L_0 := \log(\frac{\mu_0}{1-\mu_0})$ and $\ell := \log(\frac{\underline{\mu}}{1-\underline{\mu}})$, the log likelihood ratios of the initial belief μ_0 and the cutoff belief $\underline{\mu}$. Fix any $b_D > 0$ and $b_H > 0$.*

(a) If $L_0 \leq \ell - b_D$, there exists $\kappa_D > 0$ such that

$$\mathbb{P}(\lim_{t\to\infty} \mu_t = 0) \geq 1 - e^{-\kappa_D b_D} > 0.$$

(b) If $L_0 \geq \ell + b_H$, there exists $\kappa_H > 0$ such that

$$\mathbb{P}(\lim_{t\to\infty} \mu_t = 1) \geq 1 - e^{-\kappa_H b_H} > 0.$$

2. For any $\mu_0 \in (0,1)$, $\mathbb{P}(\lim_{t\to\infty} \mu_t = 0 \text{ or } \lim_{t\to\infty} \mu_t = 1) = 1$. Also, for any $\mu_0 \in (0,1)$, $\mathbb{P}(\lim_{t\to\infty} \mu_t = 0) > 0$ and $\mathbb{P}(\lim_{t\to\infty} \mu_t = 1) > 0$.

In Part 1(a), the equation $1 - e^{-\kappa_D b_D}$ bounds the probability of convergence to disenchantment ($\mu_t \to 0$). This lower bound increases with B, which measures how small the initial belief μ_0 is relative to the cutoff belief $\underline{\mu}$. Part 1(b) obtains a similar bound to the probability of convergence to full conviction ($\mu_t \to 1$), which increases with how large μ_0 is relative to $\underline{\mu}$ (i.e., b_H). The initial belief μ_0 is thus informative on long-run belief, in contrast to the irrelevance of μ_0 for strict religions.

Part 2 further shows a bifurcation of long-run belief: almost surely belief converges to either disenchantment ($\mu_t \to 0$) or conviction ($\mu_t \to 1$); the probability of convergence to either outcome is strictly positive, which generalizes Proposition B.3. This demonstrates that even the most fortunate is not a guaranteed social base for religions of world adjustment.

Why do we have the two equilibrium paths, $\mu_t \to 0$ and $\mu_t \to 1$? Intuitively, disbelief and disobedience reinforce each other for the fortunate. A disbelieving individual disobeys, so good fortunes further undermine religious belief ("a sinner is enjoying life"), forming the path $\mu_t \to 0$. World adjustment admits another path for the fortunate, since belief and obedience also reinforce each other: a believing individual obeys, so good fortunes further strengthens religious belief ("a just person is enjoying life"), forming the path $\mu_t \to 1$.

[14] The lower bound adds $\varepsilon > 0$ to avoid pathologies in belief convergence. Also, if $\alpha = \lim_{t\to\infty}(j_t/i_t)$ exists, apparently $\alpha \geq \alpha_1 + \varepsilon$. We don't need the existence of the limit, however, for all the propositions in this section. We keep $\alpha = \lim_{t\to\infty}(j_t/i_t)$ so we can compare our results with the previous section.

The problem of legibility and the divergence in church capacity. The analysis above shows that early life outcomes are informative on long-run belief of world adjustment. Early life experiences such as the *exact sequence* of early fortunes are very difficult to observe for a church. But these early experiences are informative under world adjustment because belief exhibits path dependence.

Religiosity is therefore much less legible under world adjustment than under a strict religion, which may generate a divergence in church capacity. In supply-side analyses of religions, a church is strong and capable if it can supply appealing public goods, either spiritual or material (Iannaccone (1992); Carvalho (2020)). When approached by their church, a more devoted follower usually contributes more to public goods. Because a church of a strict religion has a much better idea of the religiosity of each individual, the church is in a better position to implement "price discrimination:" the church can ask for high contributions from sufferers, while also confidently minimizing the effort to collect contributions from the fortunate. A church of world adjustment, however, cannot easily implement such calibrated price discrimination, unless it has access to fine-grained data on its followers. Under strict religions, the large and stable contributions not only attract more followers by providing better public goods, but also support missionary work in distant and unfamiliar population.

In sum, our analysis in this appendix can be considered as applied models of misspecified learning (Bohren and Hauser (2025)), which are constructed for substantive questions in the economics of religion. In particular, we identify the specific types of religious commandments that drive different learning paths. The baseline model in the text, along with their central economic implications, is thus robust to Bayesian learning. Bayesian updating also uncovers the new insight that strict religions might produce followers whose religiosity is easier to observe, which offers another perspective on the potential divergence in church capacity. In Appendix C, we further show that our key insights are robust to forward-looking behaviors.

B.5 Proofs

We first introduce notation that will be used throughout the proofs. Denote

$$L_t := \log \frac{\mu_t}{1-\mu_t}$$

as the log likelihood ratio of the religious worldview. Define four strictly positive numbers:

$$x := \log \frac{\bar{\lambda}}{\lambda_M}, \qquad y := \log \frac{1-\lambda_M}{1-\bar{\lambda}}, \qquad z := \log \frac{\lambda_M}{\underline{\lambda}}, \qquad w := \log \frac{1-\underline{\lambda}}{1-\lambda_M}.$$

Bayes' rule implies the following changes in the log likelihood ratio:

	bad fortune	good fortune
$a_t \neq a^h$	x	$-y$
$a_t = a^h$	$-z$	w.

(B.7)

In this notation,

$$\alpha_0 = \frac{x}{y} \qquad \text{and} \qquad \alpha_1 = \frac{z}{w}.$$

We will also use the following monotonicity property. For any $0 \leq \mu < \mu' \leq 1$,

$$\Delta\big(\mu'; a^h, \psi, (\bar{\lambda} - \underline{\lambda})K\big) < \Delta\big(\mu; a^h, \psi, (\bar{\lambda} - \underline{\lambda})K\big). \tag{B.8}$$

Indeed, for every disobeying action a,

$$H(a - a^h; \mu') - \mu'(\bar{\lambda} - \underline{\lambda})K \leq H(a - a^h; \mu) - \mu(\bar{\lambda} - \underline{\lambda})K - (\mu' - \mu)(\bar{\lambda} - \underline{\lambda})K.$$

The inequality follows from $\partial H / \partial \mu < 0$. The last term is strictly negative and does not depend on a, so taking the maximum over disobeying actions establishes Equation B.8.

Lemma B.1. *For any process $\theta(\omega)$ and any $\hat{\mu} \in (0,1)$, as $t \to \infty$, $\mu_t \not\to \hat{\mu}$. In other words, the belief μ_t cannot converge to an interior value in $(0,1)$.*

Proof. Suppose, to the contrary, that $\mu_t \to \widehat{\mu}$ for some $\widehat{\mu} \in (0,1)$. Then $L_t \to \log\big(\widehat{\mu}/(1-\widehat{\mu})\big)$, so $L_{t+1} - L_t \to 0$. But by Equation B.7, every change in L_t belongs to the set $\{x, -y, -z, w\}$. All four numbers are nonzero, so

$$|L_{t+1} - L_t| \geq \min\{x, y, z, w\} > 0$$

in every period. This is a contradiction. □

Proposition B.1. *Suppose that $a^h \in \mathcal{D}$. As $t \to \infty$, the following statements are true.*

1. *If $\alpha < \alpha_0$, then under any $\mu_0 \in (0,1)$ and any process $\theta(\omega) \in \Theta(\alpha)$, $\mu_t \to 1$.*
2. *If $\alpha > \alpha_0$, then under any $\mu_0 \in (0,1)$ and any process $\theta(\omega) \in \Theta(\alpha)$, $\mu_t \to 0$.*

Proof. Suppose that $a^h \in \mathcal{D}$. By definition,

$$\Delta\big(1; a^h, \psi, (\bar{\lambda} - \underline{\lambda})K\big) \geq 0.$$

For every finite t, Bayes' rule and $0 < \underline{\lambda} < \lambda_M < \bar{\lambda} < 1$ imply $\mu_{t-1} \in (0,1)$. It follows from Equation B.8 that

$$\Delta\big(\mu_{t-1}; a^h, \psi, (\bar{\lambda} - \underline{\lambda})K\big) > 0.$$

Thus, the individual never obeys the commandment. Iterating the first row of Equation B.7 gives

$$L_t = L_0 + i_t x - j_t y.$$

For every sufficiently large t, $i_t > 0$, and therefore

$$L_t = L_0 + i_t \left(x - \frac{j_t}{i_t} y \right). \tag{B.9}$$

For every $\theta(\omega) \in \Theta(\alpha)$, $i_t \to \infty$ and $j_t / i_t \to \alpha$. If $\alpha < \alpha_0 = x/y$, the term in parentheses in Equation B.9 converges to a strictly positive number. Hence $L_t \to +\infty$ and $\mu_t \to 1$. If $\alpha > \alpha_0$, the term converges to a strictly negative number. Hence $L_t \to -\infty$ and $\mu_t \to 0$. □

Lemma B.3. *For any $a^h \in \mathcal{A}$, there exists a unique $\underline{\mu} \in (0,1)$ such that*

$$\begin{cases} a_t \neq a^h & \text{if } \mu_{t-1} \leq \underline{\mu}, \\ a_t = a^h & \text{if } \mu_{t-1} > \underline{\mu}. \end{cases}$$

Proof. Because $a^h \in \mathcal{A}$,

$$\Delta\big(1; a^h, \psi, (\bar{\lambda} - \underline{\lambda})K\big) < 0.$$

At $\mu = 0$, guilt is zero, and the individual maximizes only the secular payoff. Since $a^h \neq a^m(\psi)$,

$$\Delta\big(0; a^h, \psi, (\bar{\lambda} - \underline{\lambda})K\big) = \psi F(a^m(\psi)) - C(a^m(\psi)) - [\psi F(a^h) - C(a^h)] > 0.$$

Under the curvature conditions in the baseline model, the maximizer in the definition of Δ lies in a compact set. The Maximum Theorem therefore implies that Δ is continuous in μ. It is strictly decreasing by Equation B.8, so it has a unique zero $\underline{\mu} \in (0,1)$. The stated choice of a_t follows from the tie-breaking rule that the individual disobeys when indifferent. □

Lemma B.4.

$$\alpha_0 < \frac{1 - \lambda_M}{\lambda_M} < \alpha_1.$$

Proof. For a fixed $m \in (0,1)$, define

$$f(q) := m \log \frac{m}{q} + (1 - m) \log \frac{1 - m}{1 - q}, \qquad q \in (0,1).$$

We have

$$f'(q) = \frac{q-m}{q(1-q)}.$$

Thus, f is uniquely minimized at $q = m$, where $f(m) = 0$, and $f(q) > 0$ for $q \neq m$. Set $m = \lambda_M$. At $q = \bar{\lambda}$,

$$f(\bar{\lambda}) = (1-\lambda_M)y - \lambda_M x > 0,$$

which is equivalent to $\alpha_0 = x/y < (1-\lambda_M)/\lambda_M$. At $q = \underline{\lambda}$,

$$f(\underline{\lambda}) = \lambda_M z - (1-\lambda_M)w > 0,$$

which is equivalent to $(1-\lambda_M)/\lambda_M < z/w = \alpha_1$. □

Proposition B.2. *If $\alpha < \alpha_0$, for any $\mu_0 \in (0,1)$ and $\theta(\omega) \in \Theta(\alpha)$, $\lim_{t\to\infty} \mu_t$ does not exist.*

Proof. By Lemma B.1, belief cannot converge to a value in $(0,1)$. Suppose first that $\mu_t \to 0$. By Lemma B.3, after a finite number of periods the individual always disobeys. Thus, for some finite T and every $t > T$,

$$L_t = L_T + (i_t - i_T)x - (j_t - j_T)y.$$

Since $i_t - i_T \to \infty$ and $(j_t - j_T)/(i_t - i_T) \to \alpha < \alpha_0 = x/y$, we obtain $L_t \to +\infty$ and $\mu_t \to 1$, a contradiction.

Suppose instead that $\mu_t \to 1$. After a finite number of periods the individual always obeys. The second row of Equation B.7 then gives

$$L_t = L_T - (i_t - i_T)z + (j_t - j_T)w$$

for some finite T. Because $\alpha < \alpha_0 < \alpha_1 = z/w$, $L_t \to -\infty$ and $\mu_t \to 0$, again a contradiction. Therefore, $\lim_{t\to\infty} \mu_t$ does not exist. □

Proposition B.3. *1. Suppose that $\alpha < \alpha_1$. For any $\mu_0 \in (0,1)$ and any $\theta(\omega) \in \Theta(\alpha)$, either $\lim_{t\to\infty} \mu_t$ does not exist or $\lim_{t\to\infty} \mu_t = 0$.*

Proof. Lemma B.1 rules out convergence to an interior belief. If $\mu_t \to 1$, the individual eventually always obeys by Lemma B.3. Since $\alpha < \alpha_1 = z/w$, the second row of Equation B.7 implies $L_t \to -\infty$, a contradiction. Thus, belief either has no limit or converges to zero. □

The following probability bound will be used to prove Proposition B.4. It is stated at an arbitrary finite stopping time so that it can also be applied after a finite history of fortunes.

Lemma B.5. *Recall that* $\ell := \log\left(\underline{\mu}/(1-\underline{\mu})\right)$. *Suppose that, for some* $\widehat{p} \in (0,1)$,

$$p_t \leq \widehat{p}, \qquad \widehat{p}x < (1-\widehat{p})y, \qquad \widehat{p}z < (1-\widehat{p})w \tag{B.10}$$

in every period. There exist $\kappa_D > 0$ *and* $\kappa_H > 0$ *with the following properties. For every finite stopping time* T *and every* $b > 0$,

$$\mathbb{P}\left(L_t \leq \ell \text{ for all } t \geq T \text{ and } L_t \to -\infty \mid \mathcal{F}_T\right) \geq 1 - e^{-\kappa_D b} \tag{B.11}$$

on the event $\{L_T \leq \ell - b\}$, *and*

$$\mathbb{P}\left(L_t > \ell \text{ for all } t \geq T \text{ and } L_t \to +\infty \mid \mathcal{F}_T\right) \geq 1 - e^{-\kappa_H b} \tag{B.12}$$

on the event $\{L_T \geq \ell + b\}$.

Proof. Condition on $\mathcal{F}_T$ and relabel period T as period zero. By Lemma B.3, the individual disobeys exactly when $L_t \leq \ell$ and obeys exactly when $L_t > \ell$. The first two inequalities in Equation B.10 imply that there is a unique $\kappa_D > 0$ satisfying

$$\widehat{p}e^{\kappa_D x} + (1-\widehat{p})e^{-\kappa_D y} = 1. \tag{B.13}$$

To see existence and uniqueness, the left-hand side equals one at zero, has derivative $\widehat{p}x - (1-\widehat{p})y < 0$ at zero, is strictly convex, and converges to infinity. It is also increasing in the probability of bad fortune. Hence, whenever $L_t \leq \ell$,

$$\mathbb{E}[e^{\kappa_D(L_{t+1}-L_t)} \mid \mathcal{F}_t] \leq 1.$$

Let $\tau_0 := \inf\{t \geq 0 : L_t > \ell\}$. The stopped process

$$\exp\{\kappa_D[L_{t\wedge\tau_0} - \ell]\}$$

is therefore a nonnegative supermartingale. If $L_0 \leq \ell - b$, the supermartingale inequality gives

$$\mathbb{P}(\tau_0 < \infty) \leq e^{-\kappa_D b}. \tag{B.14}$$

On $\{\tau_0 = \infty\}$, the conditional mean of every change in L_t is at most $\widehat{p}x - (1-\widehat{p})y < 0$. The changes are uniformly bounded. Let

$$M_n := \sum_{t=0}^{n-1} \left\{L_{t+1} - L_t - \mathbb{E}[L_{t+1} - L_t \mid \mathcal{F}_t]\right\}.$$

The Azuma–Hoeffding inequality, followed by the Borel–Cantelli lemma, implies

$$\frac{M_n}{n} \to 0 \quad \text{almost surely.}$$

Therefore, on $\{\tau_0 = \infty\}$,

$$L_n \le L_0 + n[\widehat{p}x - (1-\widehat{p})y] + M_n \to -\infty.$$

Together with Equation B.14, this proves Equation B.11.

For the upper region, there is similarly a unique $\kappa_H > 0$ such that

$$\widehat{p}e^{\kappa_H z} + (1-\widehat{p})e^{-\kappa_H w} = 1. \tag{B.15}$$

Let $\tau_1 := \inf\{t \ge 0 : L_t \le \ell\}$. The process

$$\exp\{-\kappa_H[L_{t\wedge\tau_1} - \ell]\}$$

is a nonnegative supermartingale. The same argument gives $\mathbb{P}(\tau_1 < \infty) \le e^{-\kappa_H b}$ when $L_0 \ge \ell + b$. On $\{\tau_1 = \infty\}$, the conditional mean of every change in L_t is at least $-\widehat{p}z + (1-\widehat{p})w > 0$. The bounded martingale-difference argument gives

$$L_n \ge L_0 + n[-\widehat{p}z + (1-\widehat{p})w] + M_n \to +\infty.$$

This proves Equation B.12. □

Proposition B.4. *1. Denote $L_0 := \log(\frac{\mu_0}{1-\mu_0})$ and $\ell := \log(\frac{\underline{\mu}}{1-\underline{\mu}})$, the log likelihood ratios of the initial belief μ_0 and the cutoff belief $\underline{\mu}$. Fix any $b_D > 0$ and $b_H > 0$.*

(a) If $L_0 \le \ell - b_D$, there exists $\kappa_D > 0$ such that

$$\mathbb{P}(\lim_{t\to\infty} \mu_t = 0) \ge 1 - e^{-\kappa_D b_D} > 0.$$

(b) If $L_0 \ge \ell + b_H$, there exists $\kappa_H > 0$ such that

$$\mathbb{P}(\lim_{t\to\infty} \mu_t = 1) \ge 1 - e^{-\kappa_H b_H} > 0.$$

2. For any $\mu_0 \in (0,1)$, $\mathbb{P}(\lim_{t\to\infty}\mu_t = 0 \text{ or } \lim_{t\to\infty}\mu_t = 1) = 1$. Also, for any $\mu_0 \in (0,1)$, $\mathbb{P}(\lim_{t\to\infty}\mu_t = 0) > 0$ and $\mathbb{P}(\lim_{t\to\infty}\mu_t = 1) > 0$.

Proof. Let

$$\widehat{p} := \frac{1}{1+\alpha_1+\varepsilon}.$$

Condition B.6 implies $p_t \le \widehat{p}$ for every t. Since $\alpha_1 + \varepsilon > \alpha_1 > \alpha_0$, we have

$$\widehat{p}x < (1-\widehat{p})y \qquad \text{and} \qquad \widehat{p}z < (1-\widehat{p})w.$$

Thus, Lemma B.5 applies. If $L_0 \le \ell - b_D$, Equation B.11 gives

$$\mathbb{P}(\mu_t \to 0) \ge 1 - e^{-\kappa_D b_D} > 0.$$

If $L_0 \ge \ell + b_H$, Equation B.12 gives

$$\mathbb{P}(\mu_t \to 1) \ge 1 - e^{-\kappa_H b_H} > 0.$$

This proves Part 1.

We next show almost-sure convergence. Fix a number $b > 0$ and let $m := \lceil b/w \rceil$. Whenever the process enters the region $L_t > \ell$, m consecutive good fortunes keep it in that region and raise its log likelihood ratio to at least $\ell + b$. The conditional probability of these good fortunes is at least $(1-\widehat{p})^m$. By Equation B.12, once the process reaches $\ell + b$, its conditional probability of never returning to $L_t \le \ell$ is at least $1 - e^{-\kappa_H b}$. Hence, at every entrance into $L_t > \ell$, the conditional probability of staying there forever is at least

$$\eta := (1-\widehat{p})^m (1 - e^{-\kappa_H b}) > 0. \tag{B.16}$$

Let

$$\sigma_1 := \inf\{t \ge 0 : L_t > \ell\},$$

and, recursively,

$$\sigma_{n+1} := \inf\{t > \sigma_n : L_{t-1} \le \ell < L_t\}.$$

Thus, σ_n is the time of the n-th entrance into $L_t > \ell$, with the convention that the infimum of the empty set is infinity. From Equation B.16,

$$\mathbb{P}(\sigma_{n+1} < \infty \mid \mathcal{F}_{\sigma_n}) \le 1 - \eta$$

on $\{\sigma_n < \infty\}$. Iterated conditioning gives

$$\mathbb{P}(\sigma_n < \infty) \le (1-\eta)^{n-1}.$$

Therefore, the process crosses the cutoff only finitely many times almost surely. It eventually remains either in $L_t \leq \ell$ or in $L_t > \ell$. The conditional-drift argument in Lemma B.5 then implies, respectively, $L_t \to -\infty$ or $L_t \to +\infty$. Thus,

$$\mathbb{P}(\mu_t \to 0 \text{ or } \mu_t \to 1) = 1.$$

It remains to show that both limits have strictly positive probability for every $\mu_0 \in (0, 1)$. Since $p_t \in (0, 1)$, every prescribed finite sequence of bad and good fortunes has strictly positive probability. Starting from $L_0 \leq \ell$, a finite sequence of bad fortunes, stopped at the first crossing, raises L_t above ℓ; starting from $L_0 > \ell$, this preliminary sequence is unnecessary. A subsequent finite sequence of good fortunes raises L_t to at least $\ell + b$. Therefore, the probability of reaching $\ell + b$ is strictly positive. Conditional on reaching it, Equation B.12 gives a strictly positive probability of convergence to one.

Similarly, starting from $L_0 > \ell$, a finite sequence of bad fortunes, stopped at the first crossing, lowers L_t to at most ℓ; starting from $L_0 \leq \ell$, this preliminary sequence is unnecessary. A subsequent finite sequence of good fortunes lowers L_t to at most $\ell - b$. Conditional on reaching this point, Equation B.11 gives a strictly positive probability of convergence to zero. Hence, for every $\mu_0 \in (0, 1)$,

$$\mathbb{P}(\mu_t \to 0) > 0 \qquad \text{and} \qquad \mathbb{P}(\mu_t \to 1) > 0.$$

□

C Religiosity of forward-looking Bayesians

C.1 The setup

We show here that the insights of Section B extend well to a forward looking individual, with a discount factor at $\delta \in [0, 1)$. Consider the following function $V(\mu; a^h)$ defined for $a^h \neq a^m(\psi)$, which will be the value function at the start of each period t, under belief $\mu_{t-1} = \mu$ and the commandment at a^h. Also, to ensure that the action space $A \subset \mathbb{R}^+$ is non-empty and compact, assume that $A = [\underline{A}, \bar{A}]$, with $\underline{A} < \underline{a}(\psi, (\bar{\lambda} - \underline{\lambda})K)$ and $\bar{A} > \bar{a}(\psi, (\bar{\lambda} - \underline{\lambda})K)$ as in Lemma B.2. Define $V(\mu; a^h)$ recursively as follows.

$$V(\mu; a^h) := \max\{V^D(\mu; a^h), V^O(\mu; a^h)\},$$

$$V^D(\mu; a^h) := \max_{a \in A} \psi F(a) - C(a) + H(a - a^h; \mu) +$$

$$[\mu\bar{\lambda}+(1-\mu)\lambda_M][-K+\delta V(\frac{\mu\bar{\lambda}}{\mu\bar{\lambda}+(1-\mu)\lambda_M})]+[1-\mu\bar{\lambda}-(1-\mu)\lambda_M]\delta V(\frac{\mu(1-\bar{\lambda})}{\mu(1-\bar{\lambda})+(1-\mu)(1-\lambda_M)});$$

$$V^O(\mu; a^h) := \psi F(a^h) - C(a^h) +$$

$$[\mu\underline{\lambda}+(1-\mu)\lambda_M][-K+\delta V(\frac{\mu\underline{\lambda}}{\mu\underline{\lambda}+(1-\mu)\lambda_M})]+[1-\mu\underline{\lambda}-(1-\mu)\lambda_M]\delta V(\frac{\mu(1-\underline{\lambda})}{\mu(1-\underline{\lambda})+(1-\mu)(1-\lambda_M)}).$$

Also, define the following net payoff

$$\bar{\Delta}(\mu; a^h) := V^D(\mu; a^h) - V^O(\mu; a^h).$$

Proposition C.1. *The functions $V(\mu; a^h)$ and $\bar{\Delta}(\mu; a^h)$ exist and are unique. The two functions are also continuous on $\mu \in [0, 1]$.*

Proofs for all results in Appendix C are in Appendix C.6. The value function $V(\mu; a^h)$ and the disobedience function $\bar{\Delta}(\mu; a^h)$ are thus well defined and well behaved. The continuity of both functions is especially important when we extend the results from the myopic case.

Before we proceed, first note that Lemma B.1 immediately applies to the forward looking case: the lemma is entirely driven by Bayesian updating and the properties of the religious and mechanical worldviews, having nothing to do with the chosen action. We still only need to consider three possibilities: $\lim_{t\to\infty} \mu_t = 0$, $\lim_{t\to\infty} \mu_t = 1$, and $\lim_{t\to\infty} \mu_t$ does not exist.

C.2 Strict religions under forward-looking Bayesians

The following proposition shows that unattainable commandments are stricter under a forward looking individual than under a myopic individual. As a consequence, the forward looking case extends the key insight that the unfortunate are the social base of strict religions, determined by the same cutoff α_0 in Equation B.4. We also now restrict our attention to commandments that are within action space $A = [\underline{A}, \bar{A}]$, and redefine $\mathcal{D} := [\underline{A}, \underline{a}(\psi, \bar{\lambda}_H K)] \cup [\bar{a}(\psi, \bar{\lambda}_H K), \bar{A}]$. Any religion $a^h \in \mathcal{D}$ is thus a strict one for a myopic individual.

Proposition C.2. *Define*

$$\bar{\mathcal{D}} := \{a : \bar{\Delta}(\mu; a) \geq 0 \text{ for any } \mu \in [0,1]\}.$$

1. *Under any $\mu_0 \in (0,1)$, the individual always disobeys the commandment a^h if and only if $a^h \in \bar{\mathcal{D}}$. Also, $\bar{\mathcal{D}} \subset \mathcal{D}$.*

2. *Suppose that $a^h \in \bar{\mathcal{D}}$. If $\alpha < \alpha_0$, then under any $\mu_0 \in (0,1)$ and any process $\theta(\omega) \in \Theta(\alpha)$, $\mu_t \to 1$.*

3. *Suppose that $a^h \in \bar{\mathcal{D}}$. If $\alpha > \alpha_0$, then under any $\mu_0 \in (0,1)$ and any process $\theta(\omega) \in \Theta(\alpha)$, $\mu_t \to 0$.*

In the proposition, $\bar{\mathcal{D}}$ is the set of commandments that induce disobedience under a forward-looking individual. $\bar{\mathcal{D}} \subset \mathcal{D}$, so a commandment must be (weakly) stricter for a forward-looking than a myopic individual. To see this, notice that a necessary condition for a commandment $a^h \in \bar{\mathcal{D}}$, $\bar{\Delta}(1; a^h) \geq 0$, is equivalent to $a^h \in \mathcal{D}$. Under $\bar{\Delta}(1; a^h) \geq 0$, a forward-looking individual disobeys even if he fully believes in the religion. But then the full belief cannot change anymore, fixing the future payoff at $\delta V(1; a^h)$. So a fully believing individual only needs to consider the within-period payoff when choosing a_t, a decision problem that is identical to that of a myopic individual. For a forward-looking individual, a strict religion also needs to command at least $a^h \leq \underline{a}(\psi, (\bar{\lambda} - \underline{\lambda})K)$ or $a^h \geq \bar{a}(\psi, (\bar{\lambda} - \underline{\lambda})K)$, as defined by Lemma B.2. Since under $a^h \in \bar{\mathcal{D}}$, the individual never obeys the commandment, Proposition B.1 carries over immediately, so that the social base of strict religions is still unfortunate individuals ($\alpha < \alpha_0$).

C.3 World adjustment under forward-looking Bayesians

This section shows that results on world adjustment in Section B are robust to forward looking payoffs. The world-adjusting religions are as defined in Section B:

$$a^h \in \mathcal{A} = \Big\{ a \geq 0 : \Delta\big(1; a, \psi, (\bar{\lambda} - \underline{\lambda})K\big) < 0 \text{ and } a \neq a^m(\psi) \Big\}.$$

It is easy to show that $a^h \in \mathcal{A}$ is equivalent to $a^h \neq a^m(\psi)$ and $\bar{\Delta}(1; a^h) < 0$. We first consider the unfortunate under world adjustment $a^h \in \mathcal{A}$. The following lemma is especially useful for the forward-looking case.

Lemma C.1. *Suppose that $a^h \in \mathcal{A}$.*

1. *There exists a $\bar{\mu} \in (0,1)$ and for all $\mu \in (\bar{\mu}, 1)$, $\bar{\Delta}(\mu; a^h) < 0$.*
2. *There exists a $\underline{\mu} \in (0,1)$ and for all $\mu \in (0, \underline{\mu})$, $\bar{\Delta}(\mu; a^h) > 0$.*

In the forward-looking case, we cannot show that the function $\bar{\Delta}(\mu; a^h)$ is monotonic on $\mu \in [0,1]$, unlike Lemma B.3 for the myopic individual. Thus, there is no single cutoff $\underline{\mu}' \in (0,1)$ such that a forward-looking individual obeys if $\mu_{t-1} > \underline{\mu}'$ and disobeys if $\mu_{t-1} \leq \underline{\mu}'$. But we can still find two cutoffs $\underline{\mu}, \bar{\mu} \in (0,1)$, such that the individual obeys if $\mu_{t-1} > \bar{\mu}$ and disobeys if $\mu_{t-1} < \underline{\mu}$. To see this, notice that for a fully disbelieving individual ($\mu = 0$), he finds it strictly optimal to disobey any commandment $a^h \neq a^m(\psi)$. In other words,

$$\bar{\Delta}(0; a^h) = \max_a [\psi F(a) - C(a)] - [\psi F(a^h) - C(a^h)] > 0$$

Also, because the religion is world adjusting, an individual with a sufficiently strong belief will obey it, so $\bar{\Delta}(1; a^h) < 0$.[15] By the continuity of $\bar{B}(\mu; a^h)$ in μ, both the lower cutoff $\underline{\mu} \in (0,1)$ and the upper cutoff $\bar{\mu} \in (0,1)$ exist. Here we abuse the notation slightly and use $\underline{\mu}$ to denote the lower cutoff in the forward-looking case, while $\underline{\mu}$ denotes the single cutoff in the myopic case.

We can now characterize the belief dynamics under world adjustment. For the unfortunate ($\alpha < \alpha_0$), the dynamic is the same as in the myopic case.

Proposition C.3. *If $\alpha < \alpha_0$, for any $\mu_0 \in (0,1)$ and $\theta(\omega) \in \Theta(\alpha)$, $\lim_{t\to\infty} \mu_t$ does not exist.*

To show Proposition C.3, we only need to rule out $\lim_{t\to\infty} \mu_t = 0$ and $\lim_{t\to\infty} \mu_t = 1$. Suppose that for an $\alpha < \alpha_0$, there exist $\mu_0 \in (0,1)$ and $\theta(\omega) \in \Theta(\alpha)$, such that $\lim_{t\to\infty} \mu_t = 0$.

[15] This argument, and all our other results and analysis, apply generically even if we assume that the individual obeys the commandment a^h with $\bar{\Delta}(1; a^h) = 0$. We can exclude the only two a^h that solve $\bar{\Delta}(1; a^h) = 0$, which is the same equation as $\Delta(1; a^h) = 0$ for a myopic individual, and all other world adjusting religions satisfy $\bar{\Delta}(1; a^h) < 0$, which are of strictly positive measure.

But in that case, after a finite number of periods, the individual must have a belief lower than $\underline{\mu} \in (0,1)$ forever. This induces permanent disobedience thereafter. But if so, an unfortunate individual should have concluded that the religion is fully convincing ($\lim_{t\to\infty} \mu_t = 1$), since he suffers a lot under permanent disobedience. This contradicts $\lim_{t\to\infty} \mu_t = 0$. A similar argument rules out $\lim_{t\to\infty} \mu_t = 1$. For an unfortunate individual who is forward looking, belief still oscillates forever under world adjustment. The same argument also applies to the following proposition. The proposition shows that the "marginally fortunate," or $\alpha_0 < \alpha < \alpha_1$, are not stable followers of world adjustment, either. The cutoff α_1 is as defined in Section B (Equation B.5).

Proposition C.4. *If $\alpha < \alpha_1$. Then for any $\mu_0 \in (0,1)$ and any $\theta(\omega) \in \Theta(\alpha)$, either $\lim_{t\to\infty} \mu_t = 0$ or $\lim_{t\to\infty} \mu_t$ does not exist.*

Now for the fortunate ($\alpha > \alpha_1$), we impose the same notion that they are fortunate not only asymptotically, but also likely so in every period. There exists an $\varepsilon > 0$,

$$\frac{1-p_t}{p_t} \geq \alpha_1 + \varepsilon \text{ for all } t.$$

The following result shows that for the fortunate, the two stochastic basins of attraction still apply to forward-looking individuals.

Proposition C.5. *Denote $L_0 := \log(\frac{\mu_0}{1-\mu_0})$, $\underline{L} := \log(\frac{\underline{\mu}}{1-\underline{\mu}})$, and $\bar{L} := \log(\frac{\bar{\mu}}{1-\bar{\mu}})$. Fix any $b_D > 0$ and $b_H > 0$.*

1. *If $L_0 \leq \underline{L} - b_D$, there exists $\kappa_D > 0$ such that*

$$\mathbb{P}(\lim_{t\to\infty} \mu_t = 0) \geq 1 - e^{-\kappa_D b_D} > 0.$$

2. *If $L_0 \geq \bar{L} + b_H$, there exists $\kappa_H > 0$ such that*

$$\mathbb{P}(\lim_{t\to\infty} \mu_t = 1) \geq 1 - e^{-\kappa_H b_H} > 0.$$

The proposition shows that under forward-looking, there are still two probabilistic equilibrium paths, driven by the one-directional belief updating in the two intervals $(\bar{\mu}, 1)$ and $(0, \underline{\mu})$ when individuals experience good fortunes. The probability of belief convergence is bounded by the same technique from martingale theory as in Section B. The bound $1-e^{-\kappa_D b_D}$ depends on the distance between the initial belief μ_0 and the lower cutoff $\underline{\mu}$ (in log-likelihood scale), and the bound $1 - e^{-\kappa_H b_H}$ depends on the distance between μ_0 and the upper cutoff

$\bar{\mu}$. The key insight of the main model is robust: early life experiences, including the initial belief and the exact sequence of early fortunes, are informative on the long-run belief.

Taken together, our analysis shows that, for forward looking individuals, their religiosity remains less legible for world adjusting religions than strict religions. The potential divergence in church capacity follows.

C.4 Further variations of forward-looking Bayesians

C.4.1 The case of full liberation ($\underline{\lambda} = 0$)

Our results are strengthened in the canonical case of $\underline{\lambda} = 0$ in the main text, where a religious worldview should strive to provide *full* liberation from suffering. As above, we can recursively define the functions $V(\mu; a^h)$ and $\bar{\Delta}(\mu; a^h)$ in a Banach space of bounded continuous functions. Furthermore, whenever an event has zero predictive probability, its probability-weighted continuation term is defined by its continuous extension. Then Proposition C.1 also applies. In particular, under a strict religion $a^h \in \bar{\mathcal{D}}$, because the individual never obeys, the belief updating process does not depend on $\underline{\lambda}$, so Proposition C.2 still applies.

Under $\underline{\lambda} = 0$, generically world-adjusting religions lose credibility ($\mu_t \to 0$) among forward-looking Bayesians. This result generalizes Proposition 2 in the main text for world adjustment. If $\alpha < \alpha_0$, then belief must eventually enter the high-belief region in which the individual obeys; once obedience occurs, any subsequent bad fortune is impossible under the religious worldview and therefore sends the posterior belief to zero. Hence $\mu_t = 0$ is reached in finite time. If $\alpha > \alpha_0$, then under the mild condition that the probability of bad fortune is uniformly bounded away from zero, the same conclusion holds almost surely: either the individual eventually stops obeying, in which case persistent disobedience drives belief to zero, or the individual obeys infinitely often, in which case a bad fortune eventually occurs while obeying with probability one and immediately collapses belief to zero. Thus, except for the knife-edge case $\alpha = \alpha_0$, full liberation strongly pushes long-run belief under world adjustment toward disenchantment, just like the baseline model in the main text.

C.4.2 Weighted updating

The results are robust to weighted belief updating, with Bayesianism as a special example. Suppose that bad-fortune signals are weighted by $\beta > 0$ and good-fortune signals by $\gamma > 0$:

$$\begin{cases} \mu_t = \frac{\mu_{t-1}\lambda_H^\beta}{\mu_{t-1}\lambda_H^\beta + (1-\mu_{t-1})\lambda_M^\beta} & \text{if } \theta_t = 1. \\ \mu_t = \frac{\mu_{t-1}(1-\lambda_H)^\gamma}{\mu_{t-1}(1-\lambda_H)^\gamma + (1-\mu_{t-1})(1-\lambda_M)^\gamma} & \text{if } \theta_t = 0. \end{cases}$$

Then the relevant likelihood-ratio cutoffs become

$$\alpha_0^{\beta,\gamma} = \frac{\beta \log(\bar{\lambda}/\lambda_M)}{\gamma \log((1-\lambda_M)/(1-\bar{\lambda}))}, \qquad \alpha_1^{\beta,\gamma} = \frac{\beta \log(\lambda_M/\underline{\lambda})}{\gamma \log((1-\underline{\lambda})/(1-\lambda_M))}.$$

Because both cutoffs are scaled by the common factor β/γ, their ranking is preserved: $\alpha_0^{\beta,\gamma} < \alpha_1^{\beta,\gamma}$. All qualitative results therefore remain unchanged after replacing α_0, α_1 by $\alpha_0^{\beta,\gamma}, \alpha_1^{\beta,\gamma}$: strict religions still attract the unfortunate, world-adjusting religions still generate oscillation and path dependence, and the forward-looking extensions continue to imply lower legibility for world adjustment than for strict religions.

C.5 Perfect world adjustment ($a^h = a^m(\psi)$)

The only world-adjusting religion that is not discussed yet is the knife-edge case of $a^h = a^m(\psi)$, which exactly sanctifies the secular optimal action. Under any $\delta \in [0,1)$, the individual then chooses $a_t = a^m(\psi)$ for every period. The case is knife-edge both mathematically and conceptually: it is the only world-adjusting religion that does not have both obey and disobey regimes. The following result is straightforward.

Proposition C.6. *Suppose that $a^h = a^m(\psi)$. Then for any $\delta \in [0,1)$, the individual chooses $a_t = a^h$ in every period. Consequently:*

1. *If $\alpha > \alpha_1$, then under any $\mu_0 \in (0,1)$ and any $\theta(\omega) \in \Theta(\alpha)$, $\mu_t \to 1$.*
2. *If $\alpha < \alpha_1$, then under any $\mu_0 \in (0,1)$ and any $\theta(\omega) \in \Theta(\alpha)$, $\mu_t \to 0$.*

Fortunate individuals ($\alpha > \alpha_1$) are the social base for the religion $a^h = a^m(\psi)$. This is a non-generic result, however, because it relies on the permanent obedience of the commandment under any μ_0 and $\theta(\omega)$. In this case, good fortunes always confirm the worldview that just action is rewarded, and bad fortunes always undermine it. The legibility of religiosity here is the same as strict religion: knowing whether α is above α_1 is enough to infer long-run religiosity. This is knife-edge, however, because any other world adjusting religion $a^h \neq a^m(\psi)$ has both an obey regime and a disobey regime. A disobey regime always exists for $a^h \neq a^m(\psi)$ because an individual with a sufficiently low belief ($\mu < \underline{\mu}$) in Lemma C.1. will still disobey a^h, even if a^h is close to $a^m(\psi)$. The case of $a^h = a^m(\psi)$ is therefore special among world-adjusting religions; apart from $a^h = a^m(\psi)$, the legibility of world-adjusting religiosity is lower.

C.6 Proofs

Throughout this subsection, retain the notation L_t, x, y, z, w from the proofs of Appendix B. Thus, under disobedience, bad fortune raises L_t by x and good fortune lowers it by y. Under obedience, bad fortune lowers L_t by z and good fortune raises it by w. Recall also that

$$\alpha_0 = \frac{x}{y} < \alpha_1 = \frac{z}{w}.$$

Proposition C.1. *The functions $V(\mu; a^h)$ and $\bar{\Delta}(\mu; a^h)$ exist and are unique. The two functions are also continuous on $\mu \in [0, 1]$.*

Proof. For a continuous function W on $[0, 1]$, denote TW as the right-hand side of the Bellman equation after every continuation value V is replaced by W. The Bayesian posterior after each action and each fortune is a continuous function of the prior belief. The immediate payoff is also continuous, and A is compact. The Maximum Theorem therefore implies that T maps continuous functions into continuous functions.

For any two continuous functions W and W', the continuation term in either the disobedience branch or the obedience branch is a probability-weighted average. Hence, for every $\mu \in [0, 1]$,

$$|(TW)(\mu) - (TW')(\mu)| \leq \delta \max_{\mu' \in [0,1]} |W(\mu') - W'(\mu')|.$$

Here we also use the fact that taking the maximum of two numbers cannot increase the largest difference between the two pairs of numbers. Taking the maximum over μ shows that T is a contraction with modulus $\delta < 1$ on the complete space of continuous functions on $[0, 1]$, equipped with the supremum norm. The Contraction Mapping Theorem produces a unique continuous fixed point V. The same operator maps bounded functions into bounded functions, and the same contraction inequality on that complete space rules out any other bounded fixed point.

After substituting this unique V into the two branches of the Bellman equation, both $V^D(\mu; a^h)$ and $V^O(\mu; a^h)$ are continuous and unique. Their difference $\bar{\Delta}(\mu; a^h)$ is therefore continuous and unique. □

Proposition C.2. *Define*

$$\bar{\mathcal{D}} := \{a : \bar{\Delta}(\mu; a) \geq 0 \text{ for any } \mu \in [0, 1]\}.$$

1. *Under any $\mu_0 \in (0, 1)$, the individual always disobeys the commandment a^h if and only if $a^h \in \bar{\mathcal{D}}$. Also, $\bar{\mathcal{D}} \subset \mathcal{D}$.*

2. *Suppose that* $a^h \in \bar{\mathcal{D}}$. *If* $\alpha < \alpha_0$, *then under any* $\mu_0 \in (0,1)$ *and any process* $\theta(\omega) \in \Theta(\alpha)$, $\mu_t \to 1$.

3. *Suppose that* $a^h \in \bar{\mathcal{D}}$. *If* $\alpha > \alpha_0$, *then under any* $\mu_0 \in (0,1)$ *and any process* $\theta(\omega) \in \Theta(\alpha)$, $\mu_t \to 0$.

Proof. Suppose first that $a^h \in \bar{\mathcal{D}}$. Then $\bar{\Delta}(\mu; a^h) \geq 0$ at every belief, so the tie-breaking rule implies that the individual always disobeys.

Conversely, suppose that $a^h \notin \bar{\mathcal{D}}$. There is a belief $\mu \in [0,1]$ at which $\bar{\Delta}(\mu; a^h) < 0$. If this belief is interior, the individual obeys when it is used as the initial belief. If it is zero or one, continuity gives a nearby interior belief at which the same strict inequality holds. Thus, it is not true that the individual always disobeys under every initial belief. This proves the first equivalence.

At $\mu = 1$, every posterior belief remains equal to one. The common continuation value therefore cancels from the difference between disobedience and obedience, and

$$\begin{aligned}\bar{\Delta}(1; a^h) = &\max_{a \in A \setminus a^m} \{\psi F(a) - C(a) + H(a - a^h; 1)\} \\ &- \{\psi F(a^h) - C(a^h) + (\bar{\lambda} - \underline{\lambda})K\}.\end{aligned}$$

This is exactly $\Delta\big(1; a^h, \psi, (\bar{\lambda} - \underline{\lambda})K\big)$. Hence $a^h \in \bar{\mathcal{D}}$ implies $a^h \in \mathcal{D}$, which proves $\bar{\mathcal{D}} \subset \mathcal{D}$.

Finally, if $a^h \in \bar{\mathcal{D}}$, the individual always disobeys, so the belief process is exactly the one in Proposition B.1. It follows that $\mu_t \to 1$ when $\alpha < \alpha_0$ and $\mu_t \to 0$ when $\alpha > \alpha_0$. □

Lemma C.1. *Suppose that* $a^h \in \mathcal{A}$.

1. *There exists a* $\bar{\mu} \in (0,1)$ *and for all* $\mu \in (\bar{\mu}, 1)$, $\bar{\Delta}(\mu; a^h) < 0$.

2. *There exists a* $\underline{\mu} \in (0,1)$ *and for all* $\mu \in (0, \underline{\mu})$, $\bar{\Delta}(\mu; a^h) > 0$.

Proof. At $\mu = 0$, both actions generate the same distribution of fortunes, both posterior beliefs remain zero, and guilt is zero. Therefore,

$$\bar{\Delta}(0; a^h) = \max_{a \in A} \{\psi F(a) - C(a)\} - \{\psi F(a^h) - C(a^h)\} > 0,$$

where the strict inequality follows from $a^h \neq a^m(\psi)$. At $\mu = 1$, the continuation value again cancels, and the definition of $\mathcal{A}$ gives

$$\bar{\Delta}(1; a^h) = \Delta\big(1; a^h, \psi, (\bar{\lambda} - \underline{\lambda})K\big) < 0.$$

Continuity of $\bar{\Delta}(\mu; a^h)$ now gives a number $\underline{\mu} \in (0,1)$ such that $\bar{\Delta}(\mu; a^h) > 0$ for every $\mu \in (0, \underline{\mu})$, and a number $\bar{\mu} \in (0,1)$ such that $\bar{\Delta}(\mu; a^h) < 0$ for every $\mu \in (\bar{\mu}, 1)$. □

Proposition C.3. *If $\alpha < \alpha_0$, for any $\mu_0 \in (0,1)$ and $\theta(\omega) \in \Theta(\alpha)$, $\lim_{t\to\infty} \mu_t$ does not exist.*

Proof. Lemma B.1 rules out convergence to an interior belief. Suppose first that $\mu_t \to 0$. By Lemma C.1, after a finite date the individual always disobeys. For some finite T and every $t > T$,

$$L_t = L_T + (i_t - i_T)x - (j_t - j_T)y.$$

Deleting a finite history does not change the limiting ratio of good to bad fortunes. Since $\alpha < \alpha_0 = x/y$, the right-hand side converges to $+\infty$. This contradicts $\mu_t \to 0$.

Suppose instead that $\mu_t \to 1$. After a finite date the individual always obeys, so

$$L_t = L_T - (i_t - i_T)z + (j_t - j_T)w.$$

Since $\alpha < \alpha_0 < \alpha_1 = z/w$, the right-hand side converges to $-\infty$. This contradicts $\mu_t \to 1$. Thus, belief has no limit. □

Proposition C.4. *If $\alpha < \alpha_1$. Then for any $\mu_0 \in (0,1)$ and any $\theta(\omega) \in \Theta(\alpha)$, either $\lim_{t\to\infty} \mu_t = 0$ or $\lim_{t\to\infty} \mu_t$ does not exist.*

Proof. Lemma B.1 rules out convergence to an interior belief. If $\mu_t \to 1$, Lemma C.1 implies that the individual eventually always obeys. The preceding log likelihood ratio calculation then gives $L_t \to -\infty$ because $\alpha < \alpha_1 = z/w$, a contradiction. Therefore, belief either converges to zero or has no limit. □

Proposition C.5. *Denote $L_0 := \log(\frac{\mu_0}{1-\mu_0})$, $\underline{L} := \log(\frac{\underline{\mu}}{1-\underline{\mu}})$, and $\bar{L} := \log(\frac{\bar{\mu}}{1-\bar{\mu}})$. Fix any $b_D > 0$ and $b_H > 0$.*

1. *If $L_0 \leq \underline{L} - b_D$, there exists $\kappa_D > 0$ such that*

$$\mathbb{P}(\lim_{t\to\infty} \mu_t = 0) \geq 1 - e^{-\kappa_D b_D} > 0.$$

2. *If $L_0 \geq \bar{L} + b_H$, there exists $\kappa_H > 0$ such that*

$$\mathbb{P}(\lim_{t\to\infty} \mu_t = 1) \geq 1 - e^{-\kappa_H b_H} > 0.$$

Proof. Let

$$\widehat{p} := \frac{1}{1 + \alpha_1 + \varepsilon}.$$

The condition on objective fortunes implies $p_t \leq \widehat{p}$ in every period. Since $\alpha_1 + \varepsilon > \alpha_1 > \alpha_0$,

$$\widehat{p}x < (1 - \widehat{p})y, \qquad \widehat{p}z < (1 - \widehat{p})w.$$

While $L_t < \underline{L}$, Lemma C.1 implies that the individual disobeys. Stop the process at the first date at which $L_t \geq \underline{L}$. The lower-region stopped-supermartingale argument in the proof of Lemma B.5 therefore applies with $\underline{L}$ in place of its cutoff. Hence there is a $\kappa_D > 0$ such that, whenever $L_0 \leq \underline{L} - b_D$,

$$\mathbb{P}\big(L_t < \underline{L} \text{ for every } t \text{ and } L_t \to -\infty\big) \geq 1 - e^{-\kappa_D b_D}.$$

On this event, $\mu_t \to 0$, which proves the first claim.

While $L_t > \bar{L}$, the individual obeys. Stop the process at the first date at which $L_t \leq \bar{L}$. The upper-region argument in the same lemma applies with $\bar{L}$ in place of its cutoff. Thus, for some $\kappa_H > 0$, whenever $L_0 \geq \bar{L} + b_H$,

$$\mathbb{P}\big(L_t > \bar{L} \text{ for every } t \text{ and } L_t \to +\infty\big) \geq 1 - e^{-\kappa_H b_H}.$$

On this event, $\mu_t \to 1$, which proves the second claim. □

Proposition C.6. *Suppose that $a^h = a^m(\psi)$. Then for any $\delta \in [0,1)$, the individual chooses $a_t = a^h$ in every period. Consequently:*

1. *If $\alpha > \alpha_1$, then under any $\mu_0 \in (0,1)$ and any $\theta(\omega) \in \Theta(\alpha)$, $\mu_t \to 1$.*
2. *If $\alpha < \alpha_1$, then under any $\mu_0 \in (0,1)$ and any $\theta(\omega) \in \Theta(\alpha)$, $\mu_t \to 0$.*

Proof. Let

$$U_m := \psi F(a^m(\psi)) - C(a^m(\psi))$$

be the maximal secular payoff, and define

$$W(\mu) := \frac{U_m - K[\mu\underline{\lambda} + (1-\mu)\lambda_M]}{1-\delta}.$$

The function W is affine. Under the individual's Bayesian predictive probabilities, the expected posterior belief after either action equals the prior belief. The expected continuation value W after either action is therefore $W(\mu)$.

Suppose that $a^h = a^m(\psi)$. The obedience branch evaluated at W is

$$U_m - K[\mu\underline{\lambda} + (1-\mu)\lambda_M] + \delta W(\mu) = W(\mu).$$

For every action $a \in A$,

$$\psi F(a) - C(a) + H(a - a^h; \mu) \leq U_m,$$

because a^h maximizes the secular payoff and $H(a - a^h; \mu) \leq 0$. The disobedience branch evaluated at W is consequently no greater than

$$U_m - K[\mu\bar{\lambda} + (1 - \mu)\lambda_M] + \delta W(\mu) \leq W(\mu).$$

The last inequality is strict for every $\mu \in (0, 1)$. Thus W is a fixed point of the Bellman operator. Proposition C.1 implies that $V = W$, and obedience is strictly optimal at every interior belief. Since Bayesian updating preserves interior beliefs at every finite date, the individual obeys in every period.

Under permanent obedience,

$$L_t = L_0 - i_t z + j_t w = L_0 + i_t \left(-z + \frac{j_t}{i_t} w \right)$$

for every sufficiently large t. For a path in $\Theta(\alpha)$, $i_t \to \infty$ and $j_t / i_t \to \alpha$. If $\alpha > \alpha_1 = z/w$, then $L_t \to +\infty$ and $\mu_t \to 1$. If $\alpha < \alpha_1$, then $L_t \to -\infty$ and $\mu_t \to 0$. □

D Religious attitudes and political domination

Religious attitude	Religious response to political persecution
Activist	Prophets denounce tyranny more fiercely.
Mystic	Buddhists/Daoists further escape.
World adjusting	Eventually, Confucians internally collapse.

In the large literature on the church and state, the church is capable of constraining the state because the church is well institutionalized, commanding a "hard" power in organizational infrastructure (Becker and Pfaff (2023)). In comparison, our model highlights the inner "ideological" power that enables religious figures to constrain political domination, especially by comparing the "ideological powers" of the three ideal-type religions.

D.1 How religious activism reacts to political persecution

D.1.1 Kings and prophets in Jewish kingdoms

The enormous potential of an activist religion to constrain the king is well illustrated by the interaction between kings and prophets in Jewish kingdoms. As emphasized by Finer (1997), Jewish kingdoms invented a new form of government, a limited monarchy constrained by (holy) law.

> [The Jewish] monarch is bound by an explicit and written law code imposed on him, coequal with his subjects, imposed from the outside. The code does not consist simply of rituals he must perform, it is a set of explicit rules in criminal, civil, family, and property jurisdiction (pp.239, Finer (1997)).

Even more significantly, a king will be sanctioned by prophets if he transgresses these constraints, constraints that the king can never change. Here arises the defining spirit of constitutionalism.

> The moment the king seemed to be breaching the Law – not just its formal prescriptions but its ethical spirit – the prophets denounced him. Thus, the wholly novel, revolutionary concept of a limited monarchy, limited not as in the rest of the Middle East by cultic or ritual obligations, but in ever more elaborate social and ethical detail, by extraneous and immutable law (pp.240, Finer (1997)).

These constitutional constraints are primarily enforced by prophets. These prophets of full conviction in Yahweh would denounce the king or even the entire population in apostasy, despite the enormous personal cost.

> The prophets were persons called – nay, compelled – to speak this word, even against their will. [...] They were feared by the people. They were also mocked by them. Some paid a terrible price for their prophesying. Isaiah was scoffed at, Jeremiah was maltreated and ended his life in exile. He speaks, too, of a bloody persecution of former prophets (pp.264 and pp.266, Finer (1997)).
>
> Jeremiah was rejected, despised, and persecuted by his fellow Judeans, many of whom regarded him as a traitor. During his lifetime he was flogged, imprisoned, and frequently in hiding. [...] Yet he cannot hold it in—Yahweh's word rages within him, and he must prophesy (pp.292-293, Hayes (2012)).

D.1.2 Religious activism and Western political authority

Persistent activism of Western religious figures is equally salient at other critical junctures in Western history, such as the reform of Pope Gregory VII that entrenched strong religious constraints on political power in the Western world. As highlighted by Fukuyama (2011),

> Gregory had a titanic and inflexible will, and was once addressed by one of his associates in the papal party as "my holy Satan." Like Martin Luther four centuries later, he had a grand vision for a reformed church and its role in society. He could not be intimidated and was willing to see the conflict with the emperor escalate into outright war. (pp.407, Fukuyama (2011)).

Though Fukuyama (2011) hints that Gregory VII's confrontational personality is an idiosyncratic quality, our analysis shows that there might be something systematic about the "titanic and inflexible will" of many Western religious figures, in contrast to Eastern ones. Indeed, it has been well argued that the religious constraints on political power were transmitted to Christianity through the Old Testament (Weber (1952); Finer (1997)). By doing so, Ancient Judaism creates the religious foundation of rule of law in the West.

> Why was it so influential [, the Jewish idea that the king must be subject to holy law]? The collapse of the imperial authority in the West left the Church the mentor of the barbarian monarchs and peoples who succeeded; and the Bible was their handbook. (pp. 272, Finer (1997)).
>
> When a Carolingian wanted a picture of how a God-directed king should behave, his attention was directed to the Old Testament and particularly to the two books of Samuel, and the two books of Kings (pp. 25, Wallace-Hadrill (1965)).
>
> [It] cannot be strongly enough stressed that it [the Vulgate Bible] was the most influential source of governmental ideas in the Middle Ages [...] The significance

> of this Biblically based standpoint was that law, as a force that regulated society, became independent of the organ that had in the first place issued it, independent of the law-giver himself. [...] This strong entrenchment of the idea of law led to the maxim of the rule of law, the idea of the *Rechtsstaat*[...] (pp.41-42, Ullmann (1975)).

D.2 How Chinese Mandarins react to political persecution

Proposition 6 shows that Mandarins cannot sustain an active attitude toward worldly actions. Suffering will eventually propel Mandarins to take a U-turn toward "non-action," the Taoist path to equanimity. This insight offers a major clue to understanding why Confucian Mandarins fail to constrain the tyranny of Chinese emperors in a persistent manner, even though Confucian canons have explicitly imposed many constraints on the emperor's power.[16] A Chinese emperor can easily induce his Mandarins to "give up" by persecuting the Mandarins, who are bound to internally collapse and embrace Taoism.[17] Eventually, under each new persecution, Mandarins escape even further from their Confucian duties to enforce the constraints on the coercive power of the emperor. This religio-cultural foundation of Chinese autocracy relies on the dynamic complementarity between Confucianism and Taoism (and later on, Buddhism), which is a central leitmotif of Chinese culture.

> Even a fervent reformist Confucian politician like Wang Anshi (1021–1086) on several occasions resigned from office, and in the end became a recluse, adopting the sobriquet "Old Man of the Mountainside" (ban shan lao ren), and taking pleasure in writing landscape poetry. Especially when "the Way was not practiced," or "the country was without the Way," when the state was in decline or in the hands of invaders, many literati intellectuals would take refuge "in lacquered gardens or on high peaks," finding solace in the Daoism of Zhuangzi and Laozi, seeking comfort in nature and pursuing the lofty goals of "oneness with the Dao" and "the realm of heaven and earth" (pp.88-89, Li (2010)).

To be clear, there are spectacular examples where Mandarins fervently denounced an emperor's transgression of Confucian principles, while disregarding the huge personal costs. The political historian Samuel Finer actually reflects on this.

> [Although Confucianism] did not rest upon a belief in the afterlife, zealous Confucianists were prepared to face the most insufferable tortures in opposing their

[16] These constraints are well articulated by, for example, Mengzi (2008); see also Lee (1992) and Tu (1993).

[17] Mandarins can also turn to another mystic religion of Buddhism, which also became popular among Mandarins after its introduction to China (Zürcher (1959)).

> principles to the wishes of the emperor, notably in the persecution of the Tung-lin Academy in the last days of the Ming Dynasty (pp. 28, Finer (1997)).

So some Confucian Mandarins did constrain the emperor as fervently as Jewish prophets. But the Confucian stratum as a whole fails to constrain the emperor in a persistent manner. This puzzling juxtaposition could be explained by our model. For a finite period of time, a Confucian Mandarin can act in a manner exactly analogous to Jewish prophets: each political persecution can propel both Confucian Mandarins and Jewish prophets to denounce tyranny even more fiercely. But a sufficiently large number of political persecutions will utterly disenchant the Mandarin, who will then turn to mystic religions. By contrast, political persecutions can never disenchant Jewish prophets who should have acted as superhuman messengers of Yahweh. Our model predicts that the resistance of Confucian Mandarins can never *persist*, bound to eventually collapse into the mystic flight of Buddhism or Taoism. The theoretical analysis is consistent with Li (2010)'s summary of the two paths for Confucian literati-intellectuals, and the dominance of the mystic path.

> What we often see [...] is either, on the one hand, "dying for the sake of humaneness, retiring for the sake of righteousness" (the sacrifice of the individual in the service of society); or, on the other hand, reclusion, flight from political struggle to the pleasures of nature. Throughout the long history of Chinese society, the latter was the most common choice (pp.88, Li (2010)).

E Status competitions and world adjustment

World affirmation and status competitions in history In Imperial China, the Confucian institution of the civil service examination was the cornerstone of the entire society (Bai and Jia (2016); Chen et al. (2020)). Commoners competed fiercely in the civil service examination, whose winners obtained the status of Confucian Mandarins. Interpreted through our model, the civil service examination induces a strong belief among commoners that they suffer because they were not yet Confucian Mandarins. From the perspective of commoners, Mandarins have attained perfect bliss through a Confucian life that is easy and relaxed.

It is intriguing that similar examinations were occasionally organized by the Chinese empire to select Buddhist and Taoist monks (pp.95, Gernet (1995); Chao (2003)), but these examinations were much less contested. Interpreted through our model, commoners were not enthusiastic about such a contest (i.e., $e_t^* = 0$) because salvation remains unattainable for monks of Buddhism or Taoism due to the very nature of salvation religions.

Status competitions and the Weberian diagnosis of modernity Figure 6 also supports Max Weber's famous diagnosis of modernization (Weber (2004a)): the waning influence of salvation religions would be accompanied by the heightened sublimation of worldly values, inducing extravagant status competitions (pp.302, Bellah (1999); pp.3-4, Tillich (2009)). This diagnosis is supported by the systematic data collected by Turchin (2023), which empirically documented that the competition over the elite status has become extremely fierce in the modern world. Indeed, when defining faith as "the state of being ultimately concerned," Tillich (2009) highlighted success in status competitions as a central object of enchantment in the modern world.

> [Consider] the ultimate concern with "success" and with social standing and economic power. It is the god of many people in the highly competitive Western culture and it does what every ultimate concern must do: it demands unconditional surrender to its laws even if the price is the sacrifice of genuine human relations, personal conviction, and creative *eros*. Its threat is social and economic defeat, and its promise – indefinite as all such promises – the fulfillment of one's being. It is the breakdown of this kind of faith which characterizes and makes religiously important most contemporary literature. Not false calculations but a misplaced faith is revealed in novels like *Point of No Return*. When fulfilled, the promise of this faith proves to be empty (pp.3-4, Tillich (2009)).

Interpreted through our model, a culture of world affirmation relies on status competitions

to protect its plausibility, at least for "commoners." As the modern culture affirms humanity ever more resolutely (a^h moves closer to a^m), commoners will indeed contest worldly trophies with increasing fervor (e_t^* increases), driven by the conviction in the ever more redemptive value of these trophies.

Status competition under guilt-free commoners and extra material benefit from elite status We modify the main model in two ways. First, the elite status confers an extra material benefit of $B \geq 0$, in addition to the opportunity to attain "salvation." Second, a commoner does not experience guilt when he disobeys the commandment a^h. Since salvation is always unattainable under the commoner status, this might relieve a commoner of a sense of guilt. Under this setup, we will see that status competition is still strongly aligned with world adjustment.

To prepare our analysis, notice that a commoner $i \in [0,1]$ chooses his action a_t^i to maximize

$$\psi F(a_t^i) - C(a_t^i) - [\mu_{t-1}\bar{\lambda}_H + (1-\mu_{t-1})\lambda_M]K.$$

The guilt-free commoner chooses $a^m(\psi)$, the optimal secular action. In turn, the commoner's effort e_t^i in the status competition maximizes the following function:

$$-\Gamma(e_t^i) + [1 - \Pi(e_t^i; e_t^{-i})]\Big\{\psi F(a^m) - C(a^m) - [\mu_{t-1}\bar{\lambda}_H + (1-\mu_{t-1})\lambda_M]K\Big\} +$$

$$\Pi(e_t^i; e_t^{-i}) \max_a \Big\{B + \psi F(a) - C(a) + H(a - a^h; \mu_{t-1}) - [\mu_{t-1}\bar{\lambda}_H \mathbf{1}\{a \neq a^h\} + (1-\mu_{t-1})\lambda_M]K\Big\}.$$

where the second line is the probability of winning the status competition ($\Pi(e_t^i; e_t^{-i})$) times the payoff from the elite status. To ensure the existence of mystic religions, assume that B is not too large:

$$B + \max_{a>0} \big\{\psi F(a) - C(a) + H(a;1)\big\} \leq \psi F\big(a^m(\psi)\big) - C\big(a^m(\psi)\big). \tag{E.1}$$

The next proposition formalizes the alignment between world adjustment and status competition under the modified setup.

Proposition E.1. *1. There exists a unique $\underline{a}^s < a^m$ and a unique $\bar{a}^s > a^m$, such that the long-run effort $e^* > 0$ if and only if $a^h \in (\underline{a}^s, \bar{a}^s)$.*

2. For $a^h \in (\underline{a}^s, \bar{a}^s)$, the long-run effort e^* strictly increases with a^h below $a^m(\psi)$ and strictly decreases with a^h above $a^m(\psi)$.

Proof. Because a commoner cannot attain salvation, suffering has a probability of $\bar{\lambda}_H$ under

the religious worldview, independently of his action. After j sufferings,

$$\frac{\mu(j)}{1-\mu(j)} = \frac{\mu_0}{1-\mu_0}\left(\frac{\bar{\lambda}_H}{\lambda_M}\right)^j,$$

so $\mu(j) \to 1$. The guilt-free commoner then chooses $a^m(\psi)$. Propositions 1 and 2 imply that a fully believing elite obeys precisely when $\underline{a}(\psi, \bar{\lambda}_H K) < a^h < \bar{a}(\psi, \bar{\lambda}_H K)$. Hence the long-run payoff gain from elite status is

$$\begin{cases} B + \psi F(a^h) - C(a^h) + \bar{\lambda}_H K - \big[\psi F(a^m(\psi)) - C(a^m(\psi))\big], & \underline{a}(\psi, \bar{\lambda}_H K) < a^h \\ & \quad < \bar{a}(\psi, \bar{\lambda}_H K), \\ B + \max\limits_{a>0}\{\psi F(a) - C(a) + H(a - a^h; 1)\} - \big[\psi F(a^m(\psi)) - C(a^m(\psi))\big], & a^h \le \underline{a}(\psi, \bar{\lambda}_H K) \text{ or} \\ & a^h \ge \bar{a}(\psi, \bar{\lambda}_H K), \end{cases}$$

At either cutoff, full-belief indifference makes the two lines equal. The maximum theorem applies because the maximand is continuous and tends to minus infinity at both action boundaries, locally uniformly in a^h. Thus the payoff gain is continuous.

Between the two cutoffs, its derivative is $\psi F'(a^h) - C'(a^h)$. Outside them, strict concavity and the endpoint derivative signs place the unique maximizer strictly between a^h and $a^m(\psi)$; the case $a^h = 0$ follows from $F(0) = -\infty$. The envelope derivative is therefore $-H'(a - a^h; 1)$ at the maximizing action. The payoff gain consequently strictly increases below $a^m(\psi)$ and strictly decreases above it. At $a^h = a^m(\psi)$ it equals $B + \bar{\lambda}_H K > 0$, while Condition E.1 makes it weakly negative at $a^h = 0$. Strict concavity of $H(\cdot; 1)$ implies $H(x; 1) \to -\infty$ as $x \to -\infty$. Together with $\psi F(a) - C(a) \to -\infty$ as $a \to \infty$, this implies, by splitting the action space at a sufficiently large constant, that the maximized payoff converges to minus infinity as $a^h \to \infty$. Continuity and strict monotonicity give unique cutoffs $\underline{a}^s \in [0, a^m(\psi))$ and $\bar{a}^s \in (a^m(\psi), \infty)$, with a positive payoff gain exactly between them.

If this gain is nonpositive, zero is the unique optimal effort. If it is positive, the commoner's payoff is strictly concave in his own effort. Along a symmetric profile, the marginal payoff from effort is positive at zero by $\Gamma'(0) = 0$ and $\partial\Pi(0; 0)/\partial e_t^i > 0$; it strictly decreases because $\partial^2\Pi/\partial(e_t^i)^2 < 0$, $\partial^2\Pi/\partial e_t^i \partial e_t^{-i} < 0$, and $\Gamma'' > 0$. It is negative for sufficiently large effort because boundedness and concavity of Π imply $\partial\Pi(e; 0)/\partial e_t^i \to 0$, while the negative cross derivative gives $0 < \partial\Pi(e; e)/\partial e_t^i \le \partial\Pi(e; 0)/\partial e_t^i$, whereas $\Gamma'(e) > 0$ for $e > 0$. Thus a unique positive symmetric equilibrium exists exactly between $\underline{a}^s$ and $\bar{a}^s$. Finally, the symmetric marginal condition strictly decreases with effort and strictly increases with the payoff gain. The implicit function theorem therefore proves Part 2, with the stated one-sided signs at the two original religious cutoffs. □

The same as the baseline model, there is a strong alignment between world adjustment and status competition. A more world-adjusting religion (a smaller $|a^h - a^m|$) still induces a greater competition into the elite status, even though the appeal of the elite status is changed by the material benefit B and the absolution of guilt for commoners.

Notice that competition for elite status can now happen under salvation religions. In this case, however, commoners are purely competing for the extra material benefit at B, not the opportunity of salvation. The extra material benefit needs to be sufficiently large to compensate for the cost of guilt experienced by an elite, as well as the secular cost of moving away from the secular optimal action a^m.