An Economic Theory of Salvation Religions*

Weijia Li† Shaoda Wang‡

May 20, 2025

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 of salvation religions, explaining why status competition is fierce where the influence of salvation religions have declined.

*We thank Jean-Paul Carvalho for insightful and valuable discussion of our paper. We have also benefited from thoughtful comments from Robert Akerlof, Samuel Bazzi, Sascha Becker, Jeanet Sinding Bentzen, Alberto Bisin, Arthur Campbell, Ernesto Dal Bo, Sebastiano Della Lena, Quoc-Anh Do, Brad Hackinen, Leander Heldring, Ruixue Jia, Erik Kimbrough, Kaveh Majlesi, Nathan Nunn, Giacomo Ponzetto, Gerard Roland, Pablo Querubin, Jared Rubin, Ferenc Szucs, Ludger Woessmann, Yang Xie, Xiaojian Zhao, and participants of Monash brownbag series, ASREC 2024 Conference at Chapman University, and 2025 NBER Conference on Economics of Culture and Institutions. 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. (2023); Carvalho et al. (2023)). 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.

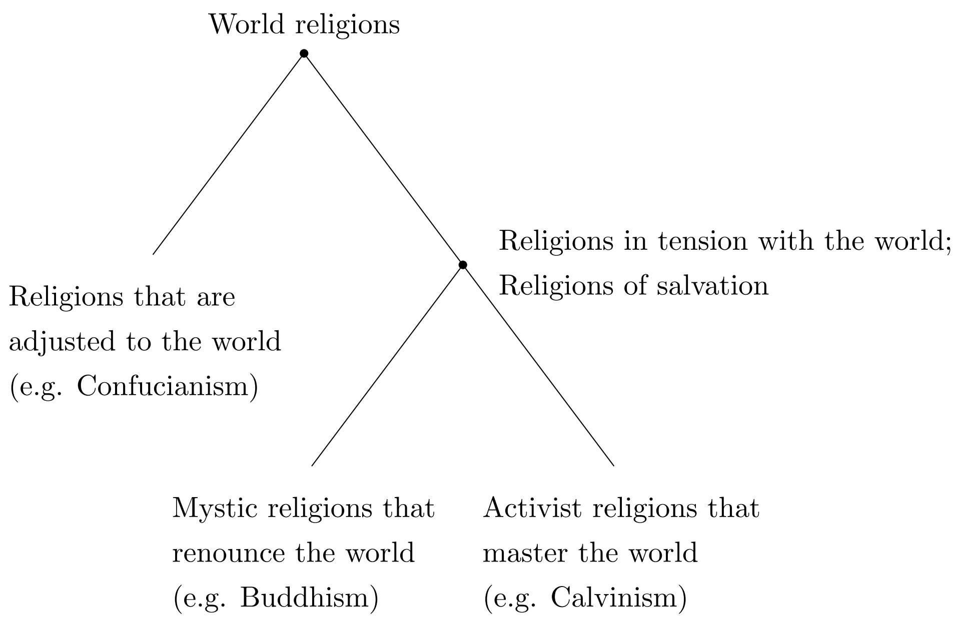


Figure 1: Weber's typology of world religions (adapted from Swedberg and Agevall (2016))

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." Foundational to this typology is a psychological definition of "salvation," the starting point of our analysis. Salvation here refers to a "permanent situation" of "an internal immunity against suffering" (pp.218, Weber (2004a)). This definition treats "salvation" as an object of social scientific study (Riesebrodt (2010); Norenzayan (2013)), which is related but different from salvation as an object of theological study (pp.66, Weber (2004b)).

Our baseline model formalizes this psychological immunity as follows. For a follower who acts as demanded by the religion, he should not psychologically suffer even when he is inflicted with disasters. But if a follower does not act as demanded, the religion warns that the follower will suffer from disasters.[1] One can immediately 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 disobeyed his God, who has imposed a just punishment on his disobedience. 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. They even transform suffering into an occasion to vitalize faith.

By contrast, if a religion imposes commandments that are too "moderate," the follower will eventually act as demanded by the religion. But 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. 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 may explain why Judeo-Christianity tended to be far more activist than salvation religions of China and India (Weber (1978); Tillich (1963)), uncovering an economic foundation for the "religious divergence" between the East and the West. 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

[1] This "commandment for salvation" has been central to social scientific analysis of salvation religions (pp.218, Weber (2004a)). Notice that even under Predestination in Calvinism, the psychological "certainty of grace" is still confirmed by human action, a famous insight from Weber (1992) and Troeltsch (1931): a Calvinist can confirm that he is actually among the "elected" by his active striving in the world.

an especially superhuman commandment because individuals must renounce the lucrative material reward from productive action. In the early Middle Ages, material rewards were indeed far more opulent in China and India (Elvin (1973); Landes (2006); Greif et al. (2020)), potentially driving the religious divergence between Eastern mysticism and Western activism.

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. (2020)). 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).

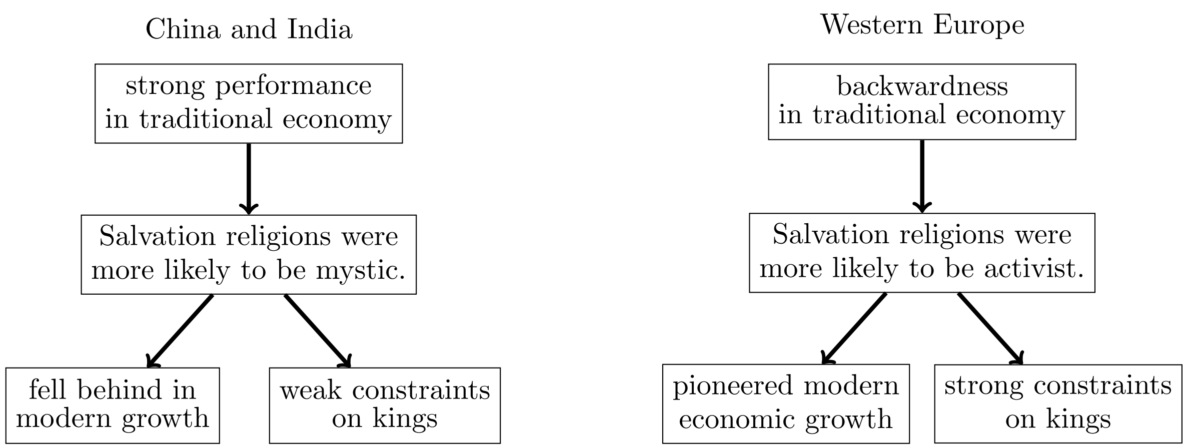


Figure 2: The potential reversal of fortune through salvation religions

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 religious figure of Judeo-Christianity may simply conclude that he suffered because he did not do enough in constraining the king (Finer (1997); Hayes (2012)). Such a belief propels an even stronger resistance against the king's domination. 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 incorporate the central leitmotif of Chinese culture, the "complementarity" between Confucianism and Taoism. Because

Confucianism is a religion of world adjustment, it cannot provide enough inner support for Chinese elites who have been persecuted by the emperor. These elites eventually 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 competitions 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 cope with 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 novel insights on how salvation religions may have contributed to the "reversal of economic fortune," complementing existing literature that pays more attention to the role of institutions and social organizations (Acemoglu et al. (2002); Greif and Tabellini (2017); Greif et al. (2020); Acemoglu and Robinson (2022); Fernández-Villaverde et al. (2023)).

Our analysis of religious coping follows the belief approach to religion (McCleary and Barro (2006); Bénabou and Tirole (2011); Levy and Razin (2012); Augenblick et al. (2016); Nunn and Sanchez de la Sierra (2017); Carvalho et al. (2023)), 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 show 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.

By focusing on religious coping with suffering, we show that religion may revitalize political resistance by endowing a special meaning with suffering in the hands of a tyrant.

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 burgeoning 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. (2024); Bisin et al. (2024)).

To our knowledge, we also offer the first model to analyze the central leitmotif of Chinese culture, the complementarity between Confucianism and Taoism (Jia and Kung (2025)). This provides a new and 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 reveals 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

$$a_t \geq 0.$$

The action a_t admits various interpretations, such as labor supply, wealth accumulation, political engagement, etc. The variable a_t in general represents the *worldly* action.[2]

For each period t, the individual may suffer from a disaster at a cost $K > 0$. Is it possible to be internally immune from suffering? More specifically, will a follower be psychologically liberated from such sufferings through "just" action? There are two competing worldviews.

[2]As a defining feature, salvation religions indeed impose strong ethical rules on everyday actions, much more so than the previous "archaic" religions that focus on sacrifice and rituals (Bellah (2011)).

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

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

The religion sanctifies an action $a^h \geq 0$ as the religious commandment, promising that the individual who chooses a^h can gain a perfect internal immunity against suffering. This perfect internal immunity ensures that an individual does not suffer even when he is inflicted with a disaster. But if the individual chooses an action different from a^h, the individual suffers with probability $\bar{\lambda}_H \in (0,1)$.[3]

Remark on the religious worldview Equation 1 captures the Weberian concept of salvation religions, which are fundamentally characterized by their "commandment for salvation" (pp.218, Weber (2004a)). "Salvation" here is psychological, referring to an "internal immunity against suffering" (pp.218, Weber (2004a); pp.66, Weber (2004b)). Such psychological immunity is supported by faith in religious salvation. Yet because psychological immunity is fundamentally *this-worldly*, it is related to but not the same as the theological state of religious salvation. Consider the "extreme" example of the Protestant doctrine of Predestination, where action cannot bring religious salvation at all. 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)).

This framework is the fundamental method of both Weber (2004a) and Tillich (1963). This framework directs our attention away from the exact theological state of religious salvation to how the quest for psychological immunity influences secular institutions (Whimster (2007)). Appendix A provides much more details regarding the Weberian definition of salvation religions.

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:

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

[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.

Under the naturalist worldview, one's own action is "meaningless" in terms of explaining or avoiding suffering. In addition, $\bar{\lambda}_H > \lambda_M$ so that a just God or cosmos inflicts pain on a disobeying individual with a probability ($\bar{\lambda}_H$) higher than 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 do so when he has just experienced good fortune. The asymmetric updating resonates with the "negativity bias," a fundamental and highly 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)).[4]

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 F.

2.1.2 The payoff function

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

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

The individual's payoff has three components.

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

1. The first component is the secular payoff. The term $v(\theta)F(a_t)$ is the material benefit, where the parameter θ determines the material return to action. The term $C(a_t)$ is the material cost of action. Assume that $v(\cdot) > 0$, $F(\cdot) > 0$, and $C(\cdot) > 0$ are all strictly increasing functions, with $F'' \leq 0$, $C'' > 0$, and $C''' \geq 0$.

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

$$H(a_t - a^h; \mu_{t-1}) = -\rho(\mu_{t-1})(a_t - a^h)^2,$$

   where $\rho(\cdot) > 0$ is an increasing function. In general, assume that $H(0; \mu_{t-1}) = 0$ and $H(x; \mu_{t-1}) < 0$ for $x \neq 0$. The guilt function $H(x; \mu_{t-1})$ increases with x for $x < 0$ and decreases with x for $x > 0$, and $\partial H/\partial \mu_{t-1} < 0$. Intuitively, the individual feels more guilt under a larger distance between his action and the commandment ($|a_t - a^h| = |x|$) or under a stronger belief in religion (μ_{t-1}). We also assume regular second and cross derivatives, as well as a boundary condition.[5]

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

3. The third component

$$-[\mu_{t-1} \cdot \mathbf{1}\{a_t \neq a^h\} \cdot \bar{\lambda}_H + (1 - \mu_{t-1}) \cdot \lambda_M]K \tag{4}$$

   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 has obeyed the commandment a^h, he does not suffer. By choosing another 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, whichever action he chooses.

   Lastly, we can 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

[5] The guilt function $H(x; \mu_{t-1})$ is strictly concave in x, and the cross derivative $\partial^2 H/\partial x \partial \mu_{t-1} > 0$ for $x < 0$ and $\partial^2 H/\partial x \partial \mu_{t-1} < 0$ for $x > 0$. We also assume that $\lim_{x \to -\infty} H'(x) \to \infty$. These conditions are quite general, which are true for $H(x; \rho) = -\rho(\mu_{t-1})x^\alpha$ under $\alpha > 1$.

direct utility from meaningfulness, however, for all our results.

Now we define another key variable:

$$a^m(\theta) = \max_x [v(\theta)F(x) - C(x)]. \tag{5}$$

The variable $a^m(\theta)$ is the **secular optimal action** that only maximizes the secular payoff. The secular optimal action $a^m(\theta)$ obviously increases with the return θ. In our model, the distance between $a^m(\theta)$ and the religious commandment a^h is of central interest.

Suffering in our model When our paper refers to suffering, it refers to pain that is brought by external disasters, or the term 4. We are **not** talking about, for example, the self-inflicted pain by choosing an action a_t away from the secular optimal action $a^m(\theta)$. 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 on 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.

The robustness of the setup The baseline model is robust to a forward looking individual (Appendix E) and full Bayesian updating (Appendix F). Because these extensions are largely technical results, we leave them to appendices.

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.[6]

The following proposition characterizes stable religions.

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

Formally, there exist a unique cutoff $\underline{a}(\theta, \bar{\lambda}_H K) < a^m(\theta)$ and a unique cutoff $\bar{a}(\theta, \bar{\lambda}_H K) > a^m(\theta)$, such that a religion a^h is stable if and only if

$$a^h < \underline{a}(\theta, \bar{\lambda}_H K) \text{ or } a^h > \bar{a}(\theta, \bar{\lambda}_H K). \tag{6}$$

Both cutoffs $\underline{a}(\theta, \bar{\lambda}_H K)$ and $\bar{a}(\theta, \bar{\lambda}_H K)$ are functions of θ, 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 $t = 1, 2, .., |a_{t+1}^* - a^h| \le |a_t^* - a^h|$, and $a_{t+1}^* \neq a^h$.

All proofs are in Appendix B. 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 $a^m(\theta)$ is large enough. ($a^h < \underline{a}(\theta, \bar{\lambda}_H K)$ or $a^h > \bar{a}(\theta, \bar{\lambda}_H K)$). Under such a large distance, it is too costly for the individual to fully obey the commandment at a^h. Instead, the individual's optimal action a_t^* always balances between the secular bliss point $a^m(\theta)$ and the commandment a^h. This is expressed by the first order condition for a_t^*:

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

The individual moves closer to a^h when he feels more guilty of disobeying the commandment of his God (a larger μ_{t-1}). But such an individual still never exactly performs a^h if the distance between a^h and $a^m(\theta)$ is sufficiently large.

[6]In principle, there can be “intermediate” religions where the belief $\mu(i)$ does not converge to the two extremes. We will see that such “intermediate” religions do not exist in our model.

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 his God was inflicting a just punishment on a "rebellious" follower. This religious explanation is more convincing than the naturalist explanation because a just God or cosmos should inflict a punishment with a higher probability on a rebellious 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 the 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 ($H = 0$), a religion become a purely intellectual exercise because even a fully believing follower still choose $a_t^* = a^m(\theta)$ 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:*

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

where the cutoffs $\underline{a}(\theta, \bar{\lambda}_H K)$ and $\bar{a}(\theta, \bar{\lambda}_H K)$ 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,

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

When a religion sanctifies an action a^h too close to the secular optimal action $a^m(\theta)$, in the long run the individual will exactly obey the religious commandment at a^h. When

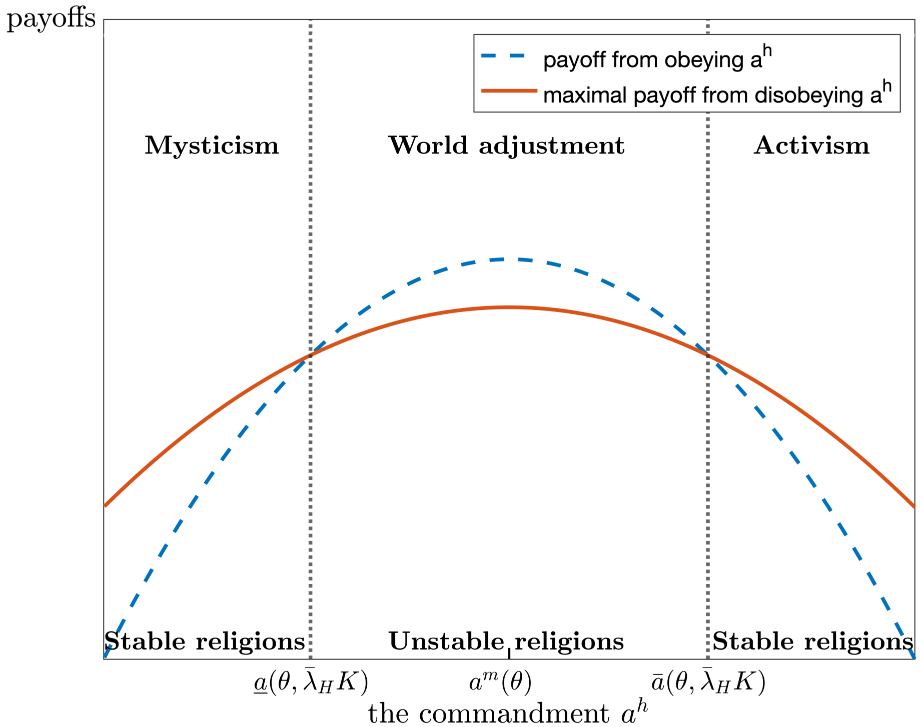


Figure 3: Religious stability and a Weberian typology of world religions

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.

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

2.3 A micro-founded Weberian typology of world religions

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

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

The corollary is also illustrated in Figure 3. Our model therefore fully 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 salvation. A salvation path is only credible if a religion is in sharp "tension with the world," or a large distance between a^h and $a^m(\theta)$. This sharp tension takes two opposite directions, activist religions that master the world ($a^h > \bar{a}(\theta, \bar{\lambda}_H K) > a^m(\theta)$) and mystic religions that renounce the world ($a^h < \underline{a}(\theta, \bar{\lambda}_H K) < a^m(\theta)$). 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 $\underline{a}(\theta, \bar{\lambda}_H K)$ and $\bar{a}(\theta, \bar{\lambda}_H K)$, which are functions of θ, 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 θ as the material motivation and $\bar{\lambda}_H K$ as the religious motivation. By doing so, the model offers a potential answer to two fundamental 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. Did salvation religions play a significant role in the great reversal of economic fortune between the East and the West?

3.1 Economic foundations of salvation religions

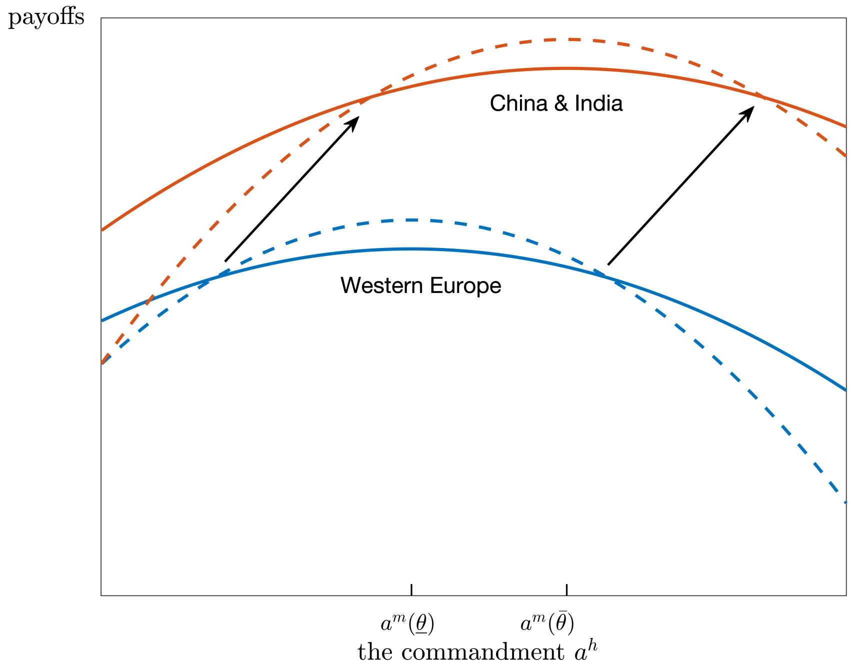
This section shows that it might be the economic backwardness of Western Europe in the early Middle Ages that drove its salvation religions to be far more activist than Chinese or Indian ones. To demonstrate this, we investigate how the material return θ affects the likelihood of different types of salvation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

Specifically, assume that the commandment a^h is a random variable. The "initial" probability density function (PDF) of a^h is $l(\cdot)$ over the support $[0, \infty)$, with the cumulative probability function $L(\cdot)$. The "final" PDF of religions for $a^h \notin [\underline{a}(\theta, \bar{\lambda}_H K), \bar{a}(\theta, \bar{\lambda}_H K)]$ is:

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

That is, for an unstable religion with

$$a^h \in \left[\underline{a}(\theta, \bar{\lambda}_H K), \bar{a}(\theta, \bar{\lambda}_H K)\right],$$



Note: the figure shows changes when the material return to action increases from $\underline{\theta}$ to $\bar{\theta} > \underline{\theta}$. 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

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 the PDF $l(a^h)$.[7] We can now characterize an economic foundation of salvation religions.

Proposition 3. *Under a higher material return θ, 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 θ drives up both cutoffs $\underline{a}(\theta, \bar{\lambda}_H K)$ and $\bar{a}(\theta, \bar{\lambda}_H K)$. Therefore, the set of mystic religions $[0, \underline{a}(\theta, \bar{\lambda}_H K))$ expands, while the set of activist religions $(\bar{a}(\theta, \bar{\lambda}_H K), \infty))$ shrinks. So we can show 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 must protect 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 ask individuals to renounce the lucrative return from worldly action. Followers of mystic

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

religions can only partially escape from lucrative action, so they can always attribute their suffering to their hesitation to detach from a world that offers such lucrative and rich benefits. This protects the promise of internal immunity 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 explain the great religious divergence between the East and the West. While activism and mysticism were both present in many salvation religions, the economic condition might determine which tendency is more dominant. Salvation religions were more mystic in China and India because they were the centers of traditional economy in the Middle Ages, allowing their elite to obtain more surplus from worldly action (Abu-Lughod (1989); Landes (2006); Greif et al. (2020)). By contrast, there were more activist denominations in Judeo-Christianity because Western Europe had long been on the economic periphery. [8]

3.2 Salvation religions can reverse "economic fortune"

Based on the logic of Proposition 3, this section shows that salvation religions might be a central mediator that reversed the "economic fortune" between the East and the West.

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:

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

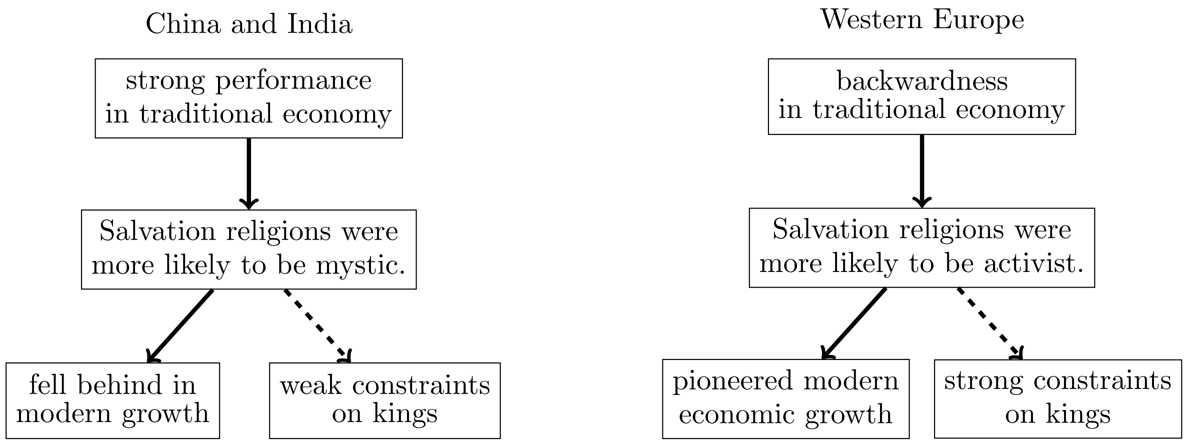
Equation 9 generates action $a^*(\theta, a^h)$ as a function of the material return θ and the commandment a^h, and a^* increases with a^h. The corresponding economic outcome

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

is therefore also a function of θ 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.

[8] Proposition 3 may also explain how a single religion evolves in its attitude towards actions. For example, early Islam was activist as "a religion of world-conquering warriors;" but as Islam conquered and controlled the economic centers of the Middle East, Sufi mysticism became prominent (pp.269, Weber (2013)).

Figure 5: The great reversal of fortune, mediated by salvation religions



Solid arrows are established by Section 3.2; dashed arrows will be established by section 4.

Specifically, we compare between two areas, the "West" and the "East," whose material returns in traditional economy are $\underline{\theta}$ and $\bar{\theta}$, with $\underline{\theta} < \bar{\theta}$. Suppose that the religious motivation is already strong enough so that religions in the West are activist and religions in the East are mystic. Denote the religion in the West as $\hat{a}^h$ and the religion in the East as $\tilde{a}^h$. We can write down their PDFs, which are assumed to be independent of each other:

$$\begin{cases} \text{West: } \hat{a}^h \sim \frac{l(x)}{1-L(\bar{a}(\underline{\theta},\bar{\lambda}_H K))} \text{ for } x \in (\bar{a}(\underline{\theta},\bar{\lambda}_H K),\bar{a}]; \\ \text{East: } \tilde{a}^h \sim \frac{l(x)}{L(\underline{a}(\bar{\theta},\bar{\lambda}_H K))} \text{ for } x \in [0,\underline{a}(\bar{\theta},\bar{\lambda}_H K)), \end{cases} \tag{11}$$

where $\bar{a} < \infty$ is the upper bound on action. Assume $Y(a^*(\bar{\theta},0)) < Y(a^*(\underline{\theta},\bar{a}))$ so that the economic outcome Y in the East under the commandment of complete inaction ($\tilde{a}^h = 0$) is lower than the economic outcome in the West under the commandment of maximal action ($\hat{a}^h = \bar{a}$). We can now establish the following result, formalizing the insight that a strong religious motivation can reverse economic fortune.

Proposition 4. *The probability that $Y(a^*(\underline{\theta},\hat{a}^h)) > Y(a^*(\bar{\theta},\tilde{a}^h))$ increases with $\bar{\lambda}_H K$. In other words, when individuals care more about religious deliverance from suffering ($\bar{\lambda}_H K$ increases), a reversal of economic fortune ($Y(a^*(\underline{\theta},\hat{a}^h)) > Y(a^*(\bar{\theta},\tilde{a}^h))$) is more likely.*

Proposition 4 shows that it is the religious motivation that drives the reversal of economic fortune. Proposition 4 and the previous Proposition 3 formally establish the solid arrows in Figure 5. Compared to the West, the East 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 East and more towards activism in the West, which tends to further reverse economic fortune.

To summarize, the reversal of economic fortune is again driven by the central principle that salvation religions tend to command the opposite of material impulse.

4 Religious tendencies and executive constraints

Another major payoff from modeling salvation religions is to shed light on a perennial puzzle in constitutionalism: how it is 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 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 meaning with suffering that revitalizes action.

4.1 Political domination and religion-motivated resistance

We embed our model of religion with 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 $Q(a_t)$, 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 $\nu(a_t)$, 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 $\kappa(a_t)$. Assume that $\nu' > 0$ and $\kappa' > 0$. If a king loses his throne, he will be replaced with another king who is *ex ante* identical.
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:

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

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

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 $C_k' > 0$, $C_k'' > 0$, and $C_k''' \geq 0$. 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 $\{1 - G[\kappa(a_t)\}$, obtaining a payoff of $E[R|R \geq \kappa(a_t)] - \kappa(a_t)$.

The king's domination D_t^* is a decreasing function of a_t, the resistance of the elite:

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

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

4.2 Activist religions constrain domination

Similar with the baseline model in Section 2, the religio-cultural elite considers two worldviews. Under each worldview, 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 the baseline model.

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

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

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 $C_e''' \geq 0$. The parameter $\omega > 0$ is the weight on the secular payoff.

2. The rest of the 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:

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

Proposition 5. *1. There exists a unique cutoff $\underline{a}(\bar{\lambda}_H K) < a^m$ and a unique cutoff $\bar{a}(\bar{\lambda}_H K) > a^m$, such that the religious worldview a^h is stable if and only if*

$$a^h < \underline{a}(\bar{\lambda}_H K) \text{ or } a^h > \bar{a}(\bar{\lambda}_H K).$$

2. Consider $a^h > \bar{a}(\bar{\lambda}_H K)$. Every period t the king persecutes the religio-cultural elite, 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.

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. 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 C elaborates on the two paradigmatic cases of religious figures in Jewish kingdoms and Western Europe of the Middle Ages.

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

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 leitmotif 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 on 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 on Confucianism, $\mu^d_{t-1} \in (0,1)$ is the belief on Taoism, and $(1 - \mu^c_{t-1} - \mu^d_{t-1})$ is the belief on 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 $\bar{\lambda}_C > \bar{\lambda}_D > \lambda_M$. 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:

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

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

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

Denote $V(a_t) = -D(a_t)Q(a_t)$ as the secular payoff from resistance. The next assumption reflects that Confucianism is world-adjusting and Taoism is mystic.

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

2. $a^d < \underline{a}(\bar{\lambda}_C K)$, $a^m < a^c < \bar{a}(\bar{\lambda}_C K)$ where $\underline{a}(\bar{\lambda}_C K)$ and $\bar{a}(\bar{\lambda}_C K)$ are 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 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 $a^c > a^m$. Only a comparison with Judeo-Christianity reveals that Confucianism is not active *enough*.

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 $\underline{\mu} \in [0, 1)$, such that for $\mu \leq \underline{\mu}$, $H(a_t - a^h; \mu) = 0$.*

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 a Mandarin starts with a sufficiently strong belief in Confucianism and a sufficiently weak belief in Taoism: $\mu_0^c > \frac{1}{2}$ and $\mu_0^d < \underline{\mu}$.*

1. *There exists an $I < \infty$, if the Mandarin is persecuted fewer than I times, the Mandarin's resistance a_t^* increases over each persecution. Therefore, the emperor's domination D_t^* decreases over each persecution.*

2. *After the $(I + 1)$-th persecution, the Mandarin's belief in Confucianism is $\mu_{t-1}^c = 0$.*

3. *If the Mandarin has been persecuted more than $I+2$ times, the Mandarin's resistance a_t^* decreases over each new persecution, driving the emperor's domination D_t^* to increase over each new persecution.*

A comparative analysis of religio-cultural elites and political domination Proposition 6 is illustrated in Figure 6. 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

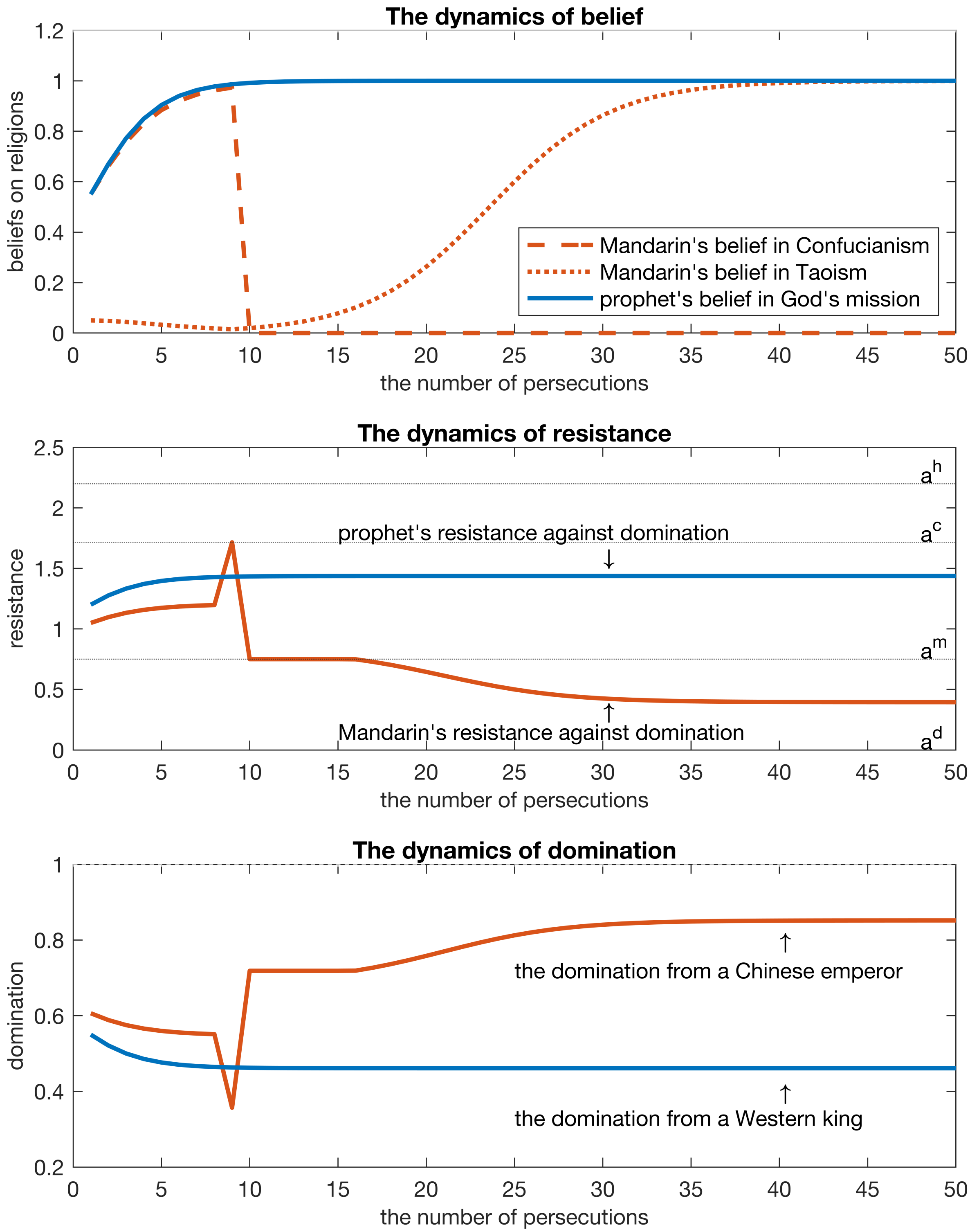


Figure 6: A comparison between a Western "prophet" and a Chinese Mandarin, as well as the domination from a Western king and a Chinese emperor

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, 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 6 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.

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 unlikely 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 $\beta > 0$ of elites and a measure 1 of commoners. At each period t, a $q > 0$ fraction of elites retire and exit the game, leaving $q\beta < 1$ elite positions available for commoners to compete. To keep the total population constant at $1 + \beta$, assume that at each period, $q\beta$ new commoners are born and enter the game.

The commoners can compete in a contest that grants elite status. Specifically, a commoner indexed as $i \in [0, 1]$ 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.

The naturalist or mechanical worldview is the same as 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."

The payoff to an individual i is as follows.

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

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 $q\beta$ measure of elites retire and exit the game.

3. Commoners chooses 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 ($\frac{\partial^2 M}{\partial (e_t^i)^2} < 0$) and an increasing marginal cost from effort ($\Gamma' > 0$, $\Gamma'' > 0$). To ensure the uniqueness of the equilibrium, we also assume the following condition.

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 is 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: $a^h \in [\underline{a}(\theta, \bar{\lambda}_H K), \bar{a}(\theta, \bar{\lambda}_H K)]$ 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 e_t^* is strictly positive after a sufficiently large number of suffering. The long-run effort strictly increases with a^h if $a^h < a^m(\theta)$ and strictly decreases with a^h if $a^h > a^m(\theta)$.

2. *Suppose that a^h is a salvation religion: $a^h \in [0, \infty) \backslash [\underline{a}(\theta, \bar{\lambda}_H K), \bar{a}(\theta, \bar{\lambda}_H K)]$ 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

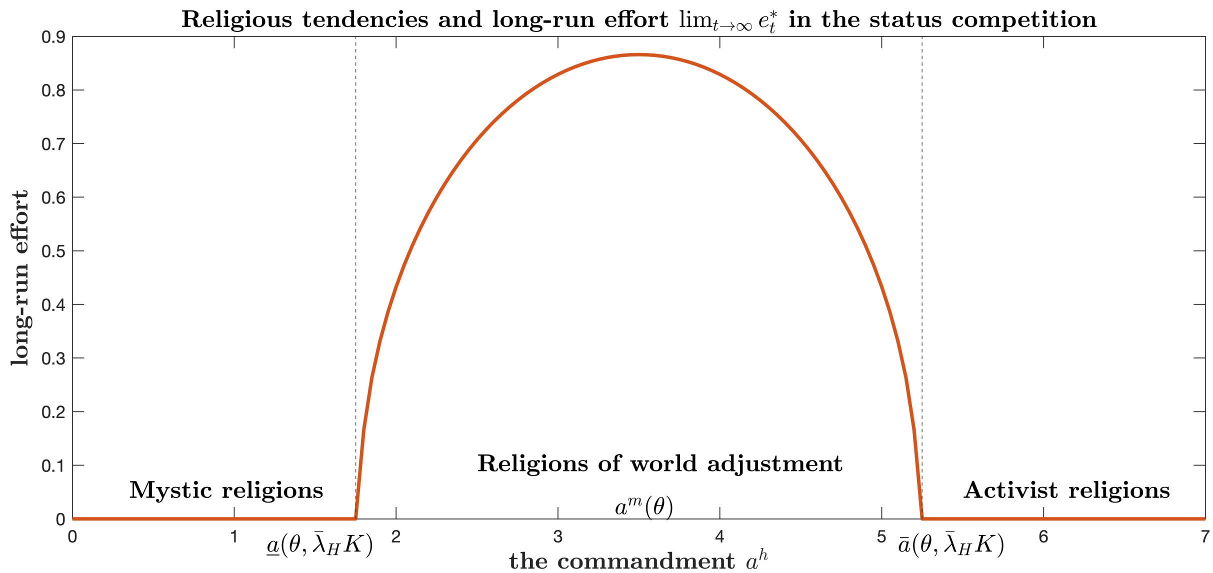


Figure 7: Religious tendencies and status competition

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 7, 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 7 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 $a^m(\theta)$), 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 discourages 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.[10] 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 7, showing an alignment between status competition and world adjusting religions. Appendix D 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 D also shows that our results are robust under modified setups. Commoners may compete for the elite status for 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).

[10] This can be easily formalized by introducing a "designer" for any religion $a^h \in [\underline{a}, \bar{a}]$ at $t = 0$, before the game starts at $t = 1$. A designer wants to maximize a function $V(x)$ that increases with x, the number of followers in the long run. The designer pays a cost D if he sets up a status competition. Suppose that $V(1) > D$, so it pays to set up the status competition if it is necessary to induce full faith among commoners. It is obvious that a designer introduces a status competition if and only if the religion is world-adjusting.

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 shed 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 "commercial 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. 2022. Weak, despotic, or inclusive? how state type emerges from state versus civil society competition. *American Political Science Review* : 1–14.

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. 2024. Religion, education, and the state. *Revise and resubmit to the Review of Economic Studies* .

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.

Becker, Sascha O., Jared Rubin, and Ludger Woessmann. 2023. Religion and Growth. *forthcoming at Journal of Economic Literature* .

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.

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

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

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.

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

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.

Blackwell, David. 1965. Discounted dynamic programming. *The Annals of Mathematical Statistics* **36**: 226–235.

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

Carvalho, Jean-Paul, Augustin Bergeron, Joseph Henrich, Nathan Nunn, and Jonathan Weigel. 2023. Zero-sum thinking, the evolution of effort-suppressing beliefs, and economic development. NBER Working Paper.

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.

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.

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. 2020. Two Paths to Prosperity: Culture and Institutions in Europe and China, 1200–2000. Book manuscript in preparation, Stanford University.

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.

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

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, and Raul Sanchez 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.

Riesebrodt, Martin. 2010. *The Promise of Salvation: A Theory of Religion*. University of Chicago Press.

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

Schluchter, Wolfgang. 1989. *Rationalism, Religion, and Domination: A Weberian Perspective*. University of California 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*. Geroge Allen & Unwin Ltd.

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

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.

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

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.) *Essential Max Weber: a Reader*. Routledge, 215–244.

Weber, Max. 2004b. Introduction the Economic Ethics of the World Religions. In Whimster, Sam (ed.) *Essential Max Weber: a Reader*. Routledge, 55–80.

Weber, Max. 2013. *From Max Weber: Essays in Sociology*. Routledge. Translated by Hans Gerth. Edited by Hans H. Gerth and C. Wright Mills.

Whimster, Sam. 2007. *Understanding Weber*. Routledge.

Zürcher, Erik. 1959. *The Buddhist Conquest of China*, volume 11. Brill Archive.

A The Weberian concept of salvation religions

Weber's definition of salvation religions Weber conceptualizes salvation religions in a critical paragraph in Weber (2004a), a foundational essay for sociology of religion.

> What was important was the content of the prophecy or the commandment for salvation: the orienting of the conduct of life for the salvation good (*Heilsgut*). This means, at least relatively speaking, the rational systematization of the conduct of life, either in its individual aspects or as a whole. The latter was the case in those religions that were specifically 'redemption' religions; that is, all those religions that offered to their followers the prospect of deliverance from suffering. This was more likely to occur, the more the nature of suffering was made more sublimated, more internalized, and more principled. What counted was to place the followers in a permanent situation that gave them an internal immunity against suffering (pp.218, Weber (2004a)).

Though quite a lot to unpack, it is clear that the "commandment for salvation" is at the heart of the Weberian concept of salvation religions, which we attempt to capture through Equation 1. "Commandment" refers to a "rational systematization of the conduct of life," either world renunciation, world adjustment, or world mastery (Schluchter (1989)).

The psychological "salvation" In "the commandment for salvation," "salvation" is eminently psychological, referring to "a permanent situation that gave them an internal immunity against suffering." This emphasis on psychological salvation is fundamental for social science of religion (Tillich (1963); Riesebrodt (2010)). Weber eloquently argued about this in another foundational essay Weber (2004b).

> In order to avoid repetition about these issues, some observations will be made in advance. Religions promise and offer different salvation goods (*Heilsgüter*), but empirical researchers do not study them only, or even mainly, as 'otherworldly'; leaving aside the fact, that by no means all religions, and also not all of the world religions, had an idea anyway of the Beyond as a site of definite promises. With the only partial exception of Christianity and a few other specifically ascetic faiths, the salvation benefits of all the religions, whether primitive or cultivated, prophetic or non-prophetic, belonged very much to this world.
>
> [...Even the] extra-world salvation goods however were by no means solely *otherworldly* (*jenseitig*), not even when they were understood by the believers to be so. For those who were seeking after salvation, it was instead to the present,

> *this-worldly* (*diesseitig*) *habitus* that they primarily turned, if one looks at this psychologically. The Puritan's certainty of salvation – the permanent state of grace that belongs to the feeling of 'affirmation' – could only be grasped psychologically through the salvation goods of this ascetic religiosity (pp.66, Weber (2004b)).

"*Habitus*" means "frame of mind, bearing, or psychological disposition" (pp.66, Weber (2004b)). Weber here put forth the key point that this psychological immunity is fundamentally this-worldly. This has been confirmed by a large empirical literature on the religious coping of disasters (Norenzayan and Hansen (2006); Hood Jr et al. (2018); Bentzen (2019); Bentzen (2021)). Therefore, "empirical researchers do not study them [salvation goods] only, or even mainly, as 'otherworldly'" (pp.66, Weber (2004b)).

As our analysis shows, this focus on "commandment for [psychological] salvation" is suitable for social science to understand how the quest for psychological immunity drives different directions of actions, deepening our understanding of how religion interacts with the secular world. It goes without saying that this Weberian conceptualization can say nothing about the theological aspect of religious salvation, which is not the domain that social science can (or should) contribute to.

It is also useful to highlight that this focus on psychological immunity is a methodological breakthrough, allowing one to study religious salvation via the conceptual tools of social sciences. The methodological point is well articulated in Whimster (2007).

> Take religious salvation, which, in non-Weberian hands, looms as a metaphysical property subject to doctrinal exegesis. With Weber, it becomes a desired end, a salvation good ('Heilsgut'); it is offered by someone (prophet or god) who is a saviour ('Heiland'); there are means for its attainment – magical rituals and ascetic or mystical ways of life ('Lebensführung'). [...] Although all these processes stem from internal mental states, they have an external empirical reference because mental states of belief are realized through actions – the cult dinner, the officiating magician, the prophet at the king's court, the prophet in the countryside excluded from court, the church priest and the institutionalization of the distribution of grace (godly favour) and, above all – and this is Weber's sociological masterstroke – the actual conduct of believers in their daily lives: conduct of life (pp.180, Whimster (2007)).

As "Weber's sociological masterstroke," he focuses on how the "internal mental states" of deliverance from sufferings interacts with "the actual conduct of believers in their daily lives: conduct of life" (pp.180, Whimster (2007)).

B Proofs for propositions in the text

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

Formally, there exist a unique cutoff $\underline{a}(\theta, \bar{\lambda}_H K) < a^m(\theta)$ and a unique cutoff $\bar{a}(\theta, \bar{\lambda}_H K) > a^m(\theta)$, such that a religion a^h is stable if and only if

$$a^h < \underline{a}(\theta, \bar{\lambda}_H K) \text{ or } a^h > \bar{a}(\theta, \bar{\lambda}_H K). \tag{6}$$

Both cutoffs $\underline{a}(\theta, \bar{\lambda}_H K)$ and $\bar{a}(\theta, \bar{\lambda}_H K)$ are functions of θ, 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 $t = 1, 2, .., |a_{t+1}^* - a^h| \leq |a_t^* - a^h|$, and $a_{t+1}^* \neq a^h$.

Proof. Define the following function:

$$B(\mu_{t-1}; \theta, \bar{\lambda}_H K) = \max_{a_t} \Big[v(\theta)F(a_t) - C(a_t) + H(a_t - a^h; \mu_{t-1}) - \mu_{t-1}\bar{\lambda}_H K\Big] - \Big[v(\theta)F(a^h) - C(a^h)\Big].$$

We first prove the following lemma:

Lemma A. *For any $\mu_{t-1} > 0$, there exists a unique $\underline{a}(\mu_{t-1}, \theta, \bar{\lambda}_H K) < a^m(\theta)$ and a unique $\bar{a}(\mu_{t-1}, \theta, \bar{\lambda}_H K) > a^m(\theta)$, such that:*

$$\begin{cases} B(\mu_{t-1}; \theta, \bar{\lambda}_H K) < 0 & \text{for all } a^h \in \Big(\underline{a}(\mu_{t-1}, \theta, \bar{\lambda}_H K), \bar{a}(\mu_{t-1}, \theta, \bar{\lambda}_H K)\Big) \\ B(\mu_{t-1}; \theta, \bar{\lambda}_H K) > 0 & \text{for all } a^h \in \Big[0, \underline{a}(\mu_{t-1}, \theta, \bar{\lambda}_H K)\Big) \cup \Big(\bar{a}(\mu_{t-1}, \theta, \bar{\lambda}_H K), \infty\Big) \\ B(\mu_{t-1}; \theta, \bar{\lambda}_H K) = 0 & \text{for } a^h = \underline{a}(\mu_{t-1}, \theta, \bar{\lambda}_H K) \text{ or } \bar{a}(\mu_{t-1}, \theta, \bar{\lambda}_H K) \end{cases}.$$

We first show the existence and uniqueness of $\bar{a}(\mu_{t-1}, \theta, \bar{\lambda}_H K) \in (a^m(\theta), \infty)$

$$\frac{\partial B}{\partial a^h} = -H'(a_t^* - a^h; \mu_{t-1}) - [v(\theta)F'(a^h) - C'(a^h)].$$

The first order condition for a_t^* is:

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

Therefore,

$$\frac{\partial B}{\partial a^h} = v(\theta)F'(a_t^*) - C'(a_t^*) - [v(\theta)F'(a^h) - C'(a^h)]$$
$$= v(\theta) \cdot \Big[F'(a_t^*) - F'(a^h)\Big] + C'(a^h) - C'(a_t^*) > 0$$

because $a_t^* < a^h$, $\frac{\partial^2 F}{\partial^2 a_t} \leq 0$, and $C'' > 0$.

Furthermore, notice that for $a^h \to a^m(\theta)$, we have $a_t^* \to a^m(\theta)$, so

$$B \to -\mu_{t-1}\bar{\lambda}_H K < 0.$$

In addition, for any $a^h > a^m(\theta)$, we have:

$$\frac{\partial B}{\partial a^h} > C'(a^h) - C'(a_t^*).$$

Because $C'' > 0$, $C''' \geq 0$, we have:

$$C'(a^h) - C'(a_t^*) \geq C''(a_t^*)(a^h - a_t^*). \tag{15}$$

Notice that $C'' > 0$ for any $a_t \geq 0$. Furthermore, as $a^h \to \infty$, $a_t^* \to \infty$, and $a^h - a_t^* \to \infty$. To see this, investigate the first order condition:

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

By contradiction, suppose that a_t^* is finite. Then as $a^h \to \infty$, $v(\theta) \cdot F'(a_t^*) - C'(a_t^*) + H'(a_t^* - a^h; \mu_{t-1}) \to \infty$, a contradiction. So $a_t^* \to \infty$.

By contradiction, suppose that $a^h - a_t^*$ is finite, so $a_t^* - a^h$ is also finite. Then as $a^h \to \infty$, $v(\theta) \cdot F'(a_t^*) - C'(a_t^*) + H'(a_t^* - a^h; \mu_{t-1}) \to -\infty$ because $C'(a_t^*) \to \infty$, a contradiction. To see that $\lim_{x\to\infty} C'(x) \to \infty$, fix an $x_0 > 0$. So for $x > x_0$, because $C''' \geq 0$, as $x \to \infty$, we have $C'(x) \geq C''(x_0)(x - x_0) \to \infty$.

Go back to Equation 15. Because $\lim_{a^h\to\infty}(a^h - a_t^*) \to \infty$, for any $\Delta > 0$, there exists $a' > 0$, such that for $a^h > a'$, we have:

$$a^h - a_t^* > \Delta.$$

Therefore, for all $a^h > a'$,

$$\frac{\partial B}{\partial a^h} \geq C''(a_t^*) \cdot \Delta > 0.$$

The function B is unbounded from the above on $a^h \in (a^m(\theta), \infty)$. Therefore $B \to \infty$ as $a^h \to \infty$. There exists a unique cutoff $\bar{a}(\theta)$ such that for $a^h > \bar{a}^h(\theta)$, $B > 0$ and for

$a^m(\theta) < a^h < \bar{a}^h(\theta)$, $B < 0$, and for $a^h = \bar{a}^h(\theta)$, $B = 0$.

We prove the existence and uniqueness of $\underline{a}(\mu_{t-1}, \theta, \bar{\lambda}_H K) \in [0, a^m(\theta))$.

First, notice that

$$\frac{\partial B(a^h)}{\partial a^h} = v(\theta) \cdot \left[F'(a_t^*) - F'(a^h)\right] + C'(a^h) - C'(a_t^*) < 0.$$

Therefore, if $B(a^h = 0) \leq 0$, assign $\underline{a}(\mu_{t-1}, \theta, \bar{\lambda}_H K) = 0$ and the claim is established. Otherwise if $B(a^h = 0) > 0$, assign $\underline{a}(\mu_{t-1}, \theta, \bar{\lambda}_H K)$ to be the unique solution to $B(a^h) = 0$ for $a^h \in (0, a^m(\theta))$. We now show that $B(a^h = 0) > 0$ is true for sufficiently large θ.

For any $a^h < a_t^* < a^m(\theta)$.

$$\frac{\partial B(a^h)}{\partial a^h} \leq C'(a^h) - C'(a_t^*).$$

Fix any $\hat{\theta}$ and $\hat{a}^h = \frac{1}{2} a^m(\hat{\theta})$. Notice that because for $\theta \to \infty$, $v(\theta) \to \infty$, so $a_t^* \to \infty$, and $C'(a_t^*) \to \infty$. For any $\Delta > 0$, there exists an $\theta'(\Delta)$ sufficiently large, such that:

$$C'(a_t^*) > (\Delta + 1) C'(\hat{a}^h) > (\Delta + 1) C'(a^h) \text{ for all } a^h < \hat{a}^h.$$

Therefore, for $\theta > \max\{\hat{\theta}, \theta'(\Delta)\}$,

$$\frac{\partial B(a^h)}{\partial a^h}\Big|_{a^h \leq \hat{a}^h} < C'(a^h) - C'(a_t^*)$$

$$< \Delta \cdot C'(a^h).$$

Therefore,

$$B(a^h = 0) = B(\hat{a}^h) - \int_0^{\hat{a}^h} \frac{\partial B(x)}{\partial x} dx$$

$$> B(\hat{a}^h) + \int_0^{\hat{a}^h} \Delta \cdot C'(x) dx$$

$$> -\mu_{t-1} \bar{\lambda}_H K + \Delta \int_0^{\hat{a}^h} C'(x) dx.$$

Take $\Delta' = \frac{\bar{\lambda}_H K}{\int_0^{\hat{a}^h} C'(x) dx} + 1$. Therefore, for $\theta > \max\{\hat{\theta}, \theta'(\Delta')\}$,

$$B(a^h = 0) > 0.$$

We have established Lemma A. Now we establish the following lemma.

Lemma B. $\bar{a}(\mu_{t-1}, \theta, \bar{\lambda}_H K)$ *increases with* μ_{t-1} *and* $\underline{a}(\mu_{t-1}, \theta, \bar{\lambda}_H K)$ *decreases with* μ_{t-1}.

The cutoff $\bar{a}(\mu_{t-1}, \theta, \bar{\lambda}_H K)$ is determined by the following equation in $\bar{a}$ and the constraint that $\bar{a} > a^m$.

$$v(\theta)F(a_t^*) - C(a_t^*) + H(a_t^* - \bar{a}; \mu_{t-1}) - \mu_{t-1}\bar{\lambda}_H K = v(\theta) \cdot F(\bar{a}) - C(\bar{a}).$$

Therefore,

$$-H'(a_t^* - \bar{a}; \mu_{t-1})d\bar{a} + \frac{\partial}{\partial \mu_{t-1}} H(a_t^* - \bar{a}; \mu_{t-1})d\mu_{t-1} - \bar{\lambda}_H K d\mu_{t-1} = [v(\theta) \cdot F'(\bar{a}) - C'(\bar{a})]d\bar{a}.$$

$$\Big\{[v(\theta) \cdot F'(\bar{a}) - C'(\bar{a})] + H'(a_t^* - \bar{a}; \mu_{t-1})\Big\}d\bar{a} = -[\bar{\lambda}_H K + \frac{\partial}{\partial \mu_{t-1}} H(a_t^* - \bar{a}; \mu_{t-1})]d\mu_{t-1},$$

$$\begin{aligned}
\frac{d\bar{a}}{d\mu_{t-1}} &= -\frac{\overbrace{\bar{\lambda}_H K - \frac{\partial}{\partial \mu_{t-1}} H(a_t^* - \bar{a}; \mu_{t-1})}^{>0}}{[v(\theta) \cdot F'(\bar{a}) - C'(\bar{a})] + H'(a_t^* - \bar{a}; \mu_{t-1})} \\
&= -\frac{\bar{\lambda}_H K - \frac{\partial}{\partial \mu_{t-1}} H(a_t^* - \bar{a}; \mu_{t-1})}{[v(\theta) \cdot F'(\bar{a}) - C'(\bar{a})] - [v(\theta)F'(a^*) - C'(a^*)]} \\
&= -\frac{\bar{\lambda}_H K - \frac{\partial}{\partial \mu_{t-1}} H(a_t^* - \bar{a}; \mu_{t-1})}{v(\theta) \cdot F'(\bar{a}) - v(\theta)F'(a^*) + C'(a^*) - C'(\bar{a})} > 0
\end{aligned}$$

because $\bar{a} > a^m(\theta)$, we have $\bar{a} > a^*$, so $v(\theta) \cdot F'(\bar{a}) - v(\theta)F'(a^*) \leq 0$ ($F'' \leq 0$) and $C'(a^*) - C'(\bar{a}) < 0$ ($C'' > 0$). Similarly, when $\underline{a} > 0$, we can show that:

$$\frac{d\underline{a}}{d\mu_{t-1}} = -\frac{\bar{\lambda}_H K - \frac{\partial}{\partial \mu_{t-1}} H(a_t^* - \underline{a}; \mu_{t-1})}{v(\theta) \cdot F'(\underline{a}) - v(\theta)F'(a^*) + C'(a^*) - C'(\underline{a})} < 0.$$

because $\underline{a} < a^m(\theta)$, we have $\underline{a} < a^*$, so $v(\theta) \cdot F'(\underline{a}) - v(\theta)F'(a^*) > 0$ ($F'' < 0$) and $C'(\underline{a}) - C'(\bar{a}) > 0$ ($C'' > 0$). For $\underline{a} = 0$, $d\underline{a}/d\mu_{t-1} = 0$ trivially. Lemma B is established.

We now prove Proposition 1.1.

Specifically, assign:

$$\begin{cases} \bar{a}(\theta, \bar{\lambda}_H K) \equiv \bar{a}(1, \theta, \bar{\lambda}_H K); \\ \bar{a}(\theta, \bar{\lambda}_H K) \equiv \underline{a}(1, \theta, \bar{\lambda}_H K). \end{cases}$$

That is, the cutoffs in Proposition 1 are the cutoffs in Lemma A for $\mu_{t-1} \to 1$. We discuss three cases.

Case 1: $a^h \in \Big[0, \underline{a}(\theta, \bar{\lambda}_H K)\Big) \cup \Big(\bar{a}(\theta, \bar{\lambda}_H K), \infty\Big)$.

By Lemma B, under any $\mu_{t-1} < 0$, for $a^h \in \Big[0, \underline{a}(\theta, \bar{\lambda}_H K)\Big)$, we must have $a^h \in \Big[0, \underline{a}(\mu_{t-1}, \theta, \bar{\lambda}_H K)\Big)$ because $\underline{a}(\mu_{t-1}, \theta, \bar{\lambda}_H K) > \underline{a}(\theta, \bar{\lambda}_H K)$. Therefore, $B(\mu_{t-1}, \theta, \bar{\lambda}_H K) > 0$.

Similarly, for $a^h \in \Big(\bar{a}(\theta, \bar{\lambda}_H K), \infty\Big)$, we must have $a^h \in \Big[\bar{a}(\mu_{t-1}, \theta, \bar{\lambda}_H K), \infty\Big)$ because $\bar{a}(\mu_{t-1}, \theta, \bar{\lambda}_H K) < \bar{a}(\theta, \bar{\lambda}_H K)$. Therefore we also have $B(\mu_{t-1}, \theta, \bar{\lambda}_H K) > 0$.

So for any $a^h \in \Big[0, \underline{a}(\theta, \bar{\lambda}_H K)\Big) \cup \Big(\bar{a}(\theta, \bar{\lambda}_H K), \infty\Big)$, $B(\mu_{t-1}, \theta, \bar{\lambda}_H K) > 0$, so the individual never chooses a^h. Instead, he chooses $a_t^* \neq a^h$ that satisfies:

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

Because the individual will never choose the commandment a^h, the belief on worldview in each period t is:

$$\mu_t = \begin{cases} \frac{\mu_{t-1}\bar{\lambda}_H}{\mu_{t-1}\bar{\lambda}_H + (1-\mu_{t-1})\lambda_M} > \mu_{t-1} & \text{iff a suffering strikes in } t \\ \mu_{t-1} & \text{iff no suffering strikes in } t \end{cases} \geq \mu_{t-1}.$$

That is, the individual's belief in the religious worldview is non-decreasing for any realized path of sufferings. For any $\epsilon > 0$, we can find a N_ϵ, such that if suffering has stricken N_ϵ times,

$$1 - \mu(N_\epsilon) < \epsilon.$$

This proves the claim.

Case 2: $a^h \in \Big[\underline{a}(\mu_0, \theta, \bar{\lambda}_H K), \underline{a}(\mu_0, \theta, \bar{\lambda}_H K)\Big]$.

Then $B(\mu_0, \theta, \bar{\lambda}_H K) < 0$. The individual chooses

$$a_t^* = a^h$$

starting from $t = 1$ until a suffering strikes. In this case, the individual believes that the probability that a suffering strikes is:

$$\mu_0 \overbrace{\lambda_H}^{=0} + (1-\mu_0)\lambda_M = (1-\mu_0)\lambda_M.$$

When the suffering strikes the first time for a period $\tau_1 \geq 1$, the individual now believes that the religious worldview that sanctifies the action a^h is false with a probability of one:

$$\mu_{\tau_1} = \frac{0}{(1-\mu_0)\lambda_M} = 0.$$

This is followed by choosing

$$a_\tau^* = \arg\max_a v(\theta)F(a) - C(a) = a^m(\theta).$$

When another suffering strikes at $\tau' > \tau$

$$\mu_{\tau'} = \frac{0}{(1-\mu_0)\lambda_M} = 0.$$

One can see for all $t \geq \tau$,

$$a_t^* = a^m(\theta) \text{ and } \mu_t = 0.$$

Case 3: { . In this case, the individual will start with choosing

$$a_1^* \neq a^h.$$

For the i-th suffering that strikes, the individual strictly increases his belief in religious worldview compared to the previous period:

$$\mu(i) = \frac{\mu(i-1)\bar{\lambda}_H}{\mu(i-1)\bar{\lambda}_H + [1-\mu(i-1)]\lambda_M} > \mu(i-1).$$

Given the assumption that $a^h \in [\underline{a}(1,\theta,\bar{\lambda}_H K), \bar{a}(1,\theta,\bar{\lambda}_H K)]$, after a sufficiently large number $I \geq 1$, we have:

$$a^h \in [\underline{a}(\mu(I),\theta,\bar{\lambda}_H K), \bar{a}(\mu(I),\theta,\bar{\lambda}_H K)].$$

Therefore, at the period $\tau(I)$ that the I-th suffering strikes, the individual chooses:

$$a_{\tau(I)}^* = a^h.$$

Then when suffering strikes at the $I+1$-th time, the individual permanently discredits the religion that sanctifies a^h:

$$\mu(I+1) = \frac{0}{[1-\mu(I)]\lambda_M} = 0.$$

We now prove Proposition 1.2.

Under $a^h < \underline{a}(\bar{\lambda}_H K)$ or $a^h > \bar{a}(\bar{\lambda}_H K)$, the individual's action a_t^* satisfies the first order condition:

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

Apply the implicit function theorem, we have:

$$\frac{da_t^*}{d\mu_{t-1}} = \frac{\frac{\partial}{\partial \mu_{t-1}} H'(a_t^* - a^h; \mu_{t-1})}{\underbrace{-\Big[v(\theta)F''(a_t^*) - C''(a_t^*) + H''(a_t^* - a^h; \mu_{t-1})\Big]}_{>0}}.$$

For $a_t^* - a^h < 0$, $\frac{\partial}{\partial \mu_{t-1}} H'(a_t^* - a^h; \mu_{t-1}) > 0$, therefore $\frac{da_t^*}{d\mu_{t-1}} > 0$; for $a_t^* - a^h > 0$, $\frac{\partial}{\partial \mu_{t-1}} H'(a_t^* - a^h; \mu_{t-1}) > 0$, therefore $\frac{da_t^*}{d\mu_{t-1}} < 0$. The claim is established. □

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

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

where the cutoffs $\underline{a}(\theta, \bar{\lambda}_H K)$ and $\bar{a}(\theta, \bar{\lambda}_H K)$ 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,

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

Proof. Proposition 2.1 is already established by the above proof for Proposition 1. Proposition 2.2 is established because for I sufficiently large, $\mu(I) = 0$. Therefore, after I sufferings,

$$\begin{aligned} a^* &= \arg\max_a v(\theta)F(a) - C(a) + H(a - a^h; 0) - \lambda_M K \\ &= \arg\max_a v(\theta)F(a) - C(a) = a^m(\theta). \end{aligned}$$

□

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

Proof. The cutoff $\bar{a}(\theta, \bar{\lambda}_H K)$ is determined by the following equation, along with the constraint that $\bar{a}(\theta, \bar{\lambda}_H K) > a^m(\theta)$:

$$\max_a \left[v(\theta)F(a) - C(a) + H(a - \bar{a}(\theta, \bar{\lambda}_H K); 1)\right] - \bar{\lambda}_H K = v(\theta) \cdot F(\bar{a}(\theta, \bar{\lambda}_H K)) - C(\bar{a}(\theta, \bar{\lambda}_H K)).$$

Apply implicit function theorem:

$$\frac{d\bar{a}(\theta, \bar{\lambda}_H K)}{d\theta} = \frac{v'(\theta)[F(\bar{a}) - F(a^*)]}{v(\theta)[F'(a^*) - F'(\bar{a})] + C'(\bar{a}) - C'(a^*)} > 0,$$

where:

$$a^* = \arg\max_a \left[v(\theta)F(a) - C(a) + H(a - \bar{a}(\theta, \bar{\lambda}_H K); 1)\right].$$

This is because $\bar{a}(\theta, \bar{\lambda}_H K) > a^* > a^m(\theta)$, so $v'(\theta)[F(\bar{a}) - F(a^*)] > 0$ and $F'(a^*) - F'(\bar{a}) \geq 0$ because $F'' \leq 0$, and $C'(\bar{a}) - C'(a^*) > 0$ because $C'' > 0$.

For the cutoff, if $\underline{a}(\theta, \bar{\lambda}_H K) = 0$, the proposition is trivially true. If $\underline{a}(\theta, \bar{\lambda}_H K) > 0$, the proof is similar as for $\bar{a}(\theta, \bar{\lambda}_H K)$. The cutoff $\underline{a}(\theta, \bar{\lambda}_H K)$ is determined by the following equation, along with the constraint that $\underline{a}(\theta, \bar{\lambda}_H K) < a^m(\theta)$:

$$\max_a \Big[v(\theta)F(a) - C(a) + H(a - \underline{a}(\theta, \bar{\lambda}_H K); 1)\Big] - \bar{\lambda}_H K = v(\theta) \cdot F(\underline{a}(\theta, \bar{\lambda}_H K)) - C(\underline{a}(\theta, \bar{\lambda}_H K)).$$

Apply implicit function theorem:

$$\frac{d\underline{a}(\theta, \bar{\lambda}_H K)}{d\theta} = \frac{v'(\theta)[F(\underline{a}) - F(a^*)]}{v(\theta)[F'(a^*) - F'(\underline{a})] + C'(\underline{a}) - C'(a^*)} > 0,$$

where:

$$a^* = \arg\max_a \Big[v(\theta)F(a) - C(a) + H(a - \underline{a}(\theta, \bar{\lambda}_H K); 1)\Big].$$

This is because $\underline{a}(\theta, \bar{\lambda}_H K) < a^* < a^m(\theta)$, so $v'(\theta)[F(\underline{a}) - F(a^*)] < 0$ and $F'(a^*) - F'(\underline{a}) \leq 0$ because $F'' \leq 0$, and $C'(\underline{a}) - C'(a^*) < 0$ because $C'' > 0$.

Finally, the probability that a salvation religion is mystic is:

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

For $L(\underline{a}(\theta, \bar{\lambda}_H K)) > 0$, the inverse of the probability is:

$$\frac{L(\underline{a}(\theta, \bar{\lambda}_H K)) + 1 - L(\bar{a}(\theta, \bar{\lambda}_H K))}{L(\underline{a}(\theta, \bar{\lambda}_H K))} = 1 + \frac{1 - L(\bar{a}(\theta, \bar{\lambda}_H K))}{L(\underline{a}(\theta, \bar{\lambda}_H K))}$$

It is easy to verify that the inverse probability decreases with θ because $\frac{\partial \underline{a}(\theta, \bar{\lambda}_H K)}{\partial \theta} \geq 0$ and $\frac{\partial \bar{a}(\theta, \bar{\lambda}_H K)}{\partial \theta} > 0$.

Similarly, the probability that a salvation religion is activist is:

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

The inverse of the probability is:

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

It is easy to verify that the inverse probability increases with θ because $\frac{\partial \underline{a}(\theta, \bar{\lambda}_H K)}{\partial \theta} \geq 0$ and $\frac{\partial \bar{a}(\theta, \bar{\lambda}_H K)}{\partial \theta} > 0$. $\square$

Proposition 4. *The probability that $Y(a^*(\underline{\theta}, \hat{a}^h)) > Y(a^*(\bar{\theta}, \tilde{a}^h))$ increases with $\bar{\lambda}_H K$. In* *other words, when individuals care more about religious deliverance from suffering ($\bar{\lambda}_H K$ increases), a reversal of economic fortune ($Y(a^*(\underline{\theta}, \hat{a}^h)) > Y(a^*(\bar{\theta}, \tilde{a}^h))$) is more likely.*

Proof. Denote:

$$Y(a^*(\underline{\theta}, \hat{a}^h)) \equiv M(\hat{a}^h) \text{ and } Y(a^*(\bar{\theta}, \tilde{a}^h)) \equiv N(\tilde{a}^h).$$

We have $M' > 0$ and $N' > 0$. The CDFs of the outcome variable Y in the West and the East are:

$$\begin{cases} Y(a^*(\underline{\theta}, \hat{a}^h)) \sim \frac{L(M^{-1}(y)) - L(M(\bar{a}(\underline{\theta}, \bar{\lambda}_H K))}{1 - L(\bar{a}(\underline{\theta}, \bar{\lambda}_H K))} \text{ for } y \in (M(\bar{a}(\underline{\theta}, \bar{\lambda}_H K)), M(\bar{a})]; \\ Y(a^*(\bar{\theta}, \tilde{a}^h)) \sim \frac{L(N^{-1}(y))}{L(\underline{a}(\bar{\theta}, \bar{\lambda}_H K))} \text{ for } y \in [N(0), N(\underline{a}(\bar{\theta}, \bar{\lambda}_H K))). \end{cases}$$

We can write down:

$$P(Y(a^*(\underline{\theta}, \hat{a}^h)) > Y(a^*(\bar{\theta}, \tilde{a}^h))) = \int_{N(0)}^{N(\underline{a}(\bar{\theta}))} \int_{y}^{M(\bar{a})} \frac{l(M^{-1}(x))M^{-1'}(x)}{1 - L(\bar{a}(\underline{\theta}))} \frac{l(N^{-1}(y))N^{-1'}(y)}{L(\underline{a}(\bar{\theta}))} dxdy.$$

$$= \int_{N(0)}^{N(\underline{a}(\bar{\theta}))} \frac{l(N^{-1}(y))N^{-1'}(y)}{L(\underline{a}(\bar{\theta}))} \left[\int_{y}^{M(\bar{a})} \frac{l(M^{-1}(x))M^{-1'}(x)}{1 - M(\bar{a}(\underline{\theta}))} dx \right] dy$$

$$= \int_{N(0)}^{N(\underline{a}(\bar{\theta}))} \frac{l(N^{-1}(y))N^{-1'}(y)}{L(\underline{a}(\bar{\theta}))} \frac{1 - L(M^{-1}(y))}{1 - L(\bar{a}(\underline{\theta}))} dy.$$

Denote $\bar{a}(\underline{\theta}) = b$ and $\underline{a}(\bar{\theta}) = a$. Further denote

$$\frac{1}{1 - L(b)} = \Pi_1,$$

$$\int_{N(0)}^{N(a)} \frac{l(N^{-1}(y))N^{-1'}(y)[1 - L(M^{-1}(y))]}{L(a)} dy = \Pi_2.$$

Therefore,

$$P(Y(a^*(\underline{\theta}, \hat{a}^h)) > Y(a^*(\bar{\theta}, \tilde{a}^h))) = \Pi_1 \cdot \Pi_2.$$

Notice that $d\Pi_1/db > 0$. We aim to prove that $db/d\bar{\lambda}_H K > 0$, $da/d\bar{\lambda}_H K < 0$, and $d\Pi_2/da < 0$, which will establish that $P(Y(a^*(\underline{\theta}, a^h)) > Y(a^*(\bar{\theta}, a^h)))$ increases with $\bar{\lambda}_H K$ because:

$$\frac{dP(Y(a^*(\underline{\theta}, \hat{a}^h)) > Y(a^*(\bar{\theta}, \tilde{a}^h)))}{d\bar{\lambda}_H K} = \Pi_2 \frac{d\Pi_1}{db} \frac{db}{d\bar{\lambda}_H K} + \Pi_1 \frac{d\Pi_2}{da} \frac{da}{d\bar{\lambda}_H K}.$$

1. We show that $db/d\bar{\lambda}_H K > 0$ and $da/d\bar{\lambda}_H K < 0$.

The variable b is determined by the following equation, with the constraint that $b > a^m =$

$\arg\max_x[v(\theta)F(x) - C(x)]$:

$$v(\underline{\theta})F(a^*) - C(a^*) + H(a^* - b; 1) - \bar{\lambda}_H K = v(\underline{\theta})L(b) - C(b).$$

The derivative $db/d\bar{\lambda}_H K$ is:

$$\frac{db}{d\bar{\lambda}_H K} = \frac{1}{-H'(a^* - b; 1) - [v(\underline{\theta})F'(b) - C'(b)]}.$$

From the first order condition for a^*:

$$v(\underline{\theta})F'(a^*) - C'(a^*) + H'(a^* - b; 1) = 0.$$

Substitute $-H'(a^* - b; 1)$ with $v(\theta)F'(a^*) - C'(a^*)$, we have:

$$\frac{db}{d\bar{\lambda}_H K} = \frac{1}{v(\underline{\theta})F'(a^*) - C'(a^*) - [v(\underline{\theta})F'(b) - C'(b)]}$$

$$= \frac{1}{v(\underline{\theta})[F'(a^*) - F'(b)] + C'(b) - C'(a^*)} > 0$$

because $b > a^* > a^m$, so $F'(a^*) - F'(b) \geq 0$ and $C'(b) - C'(a^*) > 0$. Similarly,

$$\frac{da}{d\bar{\lambda}_H K} = \frac{1}{v(\bar{\theta})[F'(a^*) - F'(a)] + C'(a) - C'(a^*)} < 0$$

because $a < a^* < a^m$, so $F'(a^*) - F'(a) < 0$ and $C'(a) - C'(a^*) < 0$.

2. We prove that $d\Pi_2/da < 0$.

$$\frac{d\Pi_2}{da} = \frac{d}{da}\int_{N(0)}^{N(a)} \frac{l(N^{-1}(y))N^{-1\prime}(y)[1 - M(N^{-1}(y))]}{L(a)}dy$$

$$= \frac{1}{L(a)^2}\Bigg\{l(N^{-1}(N(a)))N^{-1\prime}(N(a))[1 - L(M^{-1}(N(a)))]N'(a) \cdot L(a)$$

$$- \int_{N(0)}^{N(a)} l(N^{-1}(y))N^{-1\prime}(y)[1 - L(M^{-1}(y))]dy \cdot l(a)\Bigg\}$$

Notice that $N^{-1}(N(a)) = a$ and

$$N^{-1\prime}(N(a)) = \frac{1}{N'(N^{-1}(N(a)))} = \frac{1}{N'(a)},$$

we have:

$$\frac{d\Pi_2}{da} = \frac{l(a)}{L(a)^2}\left\{[1 - L(M^{-1}(N(a)))]L(a) - \int_{N(0)}^{N(a)} l(N^{-1}(y))N^{-1'}(y)[1 - L(M^{-1}(y))]dy\right\}.$$

Also, notice that

$$L(a) = L(N^{-1}(N(a))) = \int_{N(0)}^{N(a)} l(N^{-1}(y))N^{-1}(y))dy.$$

Therefore,

$$\begin{aligned}\frac{d\Pi_2}{da} &= \frac{l(a)}{L(a)^2}\left\{\int_{N(0)}^{N(a)} [1 - L(M^{-1}(N(a)))]l(N^{-1}(y))N^{-1}(y))dy\right.\\ &\quad \left. - \int_{N(0)}^{N(a)} l(N^{-1}(y))N^{-1'}(y)[1 - L(M^{-1}(y))]dy\right\}\\ &= \frac{l(a)}{L(a)^2}\int_{N(0)}^{N(a)} [L(M^{-1}(y)) - L(M^{-1}(N(a)))]dy < 0.\end{aligned}$$

□

Proposition 5. *1. There exists a unique cutoff $\underline{a}(\bar{\lambda}_H K) < a^m$ and a unique cutoff $\bar{a}(\bar{\lambda}_H K) > a^m$, such that the religious worldview a^h is stable if and only if*

$$a^h < \underline{a}(\bar{\lambda}_H K) \text{ or } a^h > \bar{a}(\bar{\lambda}_H K).$$

2. Consider $a^h > \bar{a}(\bar{\lambda}_H K)$. Every period t the king persecutes the religio-cultural elite, 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. Denote $V(a_t) \equiv -D(a_t)Q(a_t)$. Because $D'(a_t) < 0$, $Q'(a_t) < 0$, $D''(a_t) \leq 0$, and $Q''(a_t) \leq 0$, we have $V'(a_t) = -D'(a_t)Q(a_t) - D(a_t)Q'(a_t) > 0$ and $V''(a_t) = -D''(a_t)Q(a_t) - D(a_t)Q''(a_t) - 2D'(a_t)Q'(a_t) < 0$. The proof for $\bar{a}(\bar{\lambda}_H K) \in (a^m, \infty)$ and $\underline{a}(\bar{\lambda}_H K) \in [0, a^m)$ follows the same strategy as Proposition 1. We now show a sufficient condition for $B(a^h = 0) > 0$: under $a_t \in [0, \infty)$, ω sufficiently large and η sufficiently small, with $\lim_{a\to\infty} V'(a) = 0$, the same as $\lim_{a\to\infty} F'(a) = 0$ for the baseline model. For notational simplicity also ignore the subscript e in the secular cost function, so that $C_e(a_t) = C(a_t)$. Notice that:

$$\frac{\partial B(a^h)}{\partial a^h} = \omega\Big[V'(a^*_t) - V'(a^h)\Big] + \omega\eta\Big[C'(a^h) - C'(a^*_t)\Big] < 0.$$

Furthermore, for any $a^h < a_t^* < a^m$.

$$\frac{\partial B(a^h)}{\partial a^h} < \omega \cdot \Big[V'(a_t^*) - V'(a^h) \Big].$$

Fix any $\hat{\eta}$ and $\hat{\omega}$ and an $\hat{a}^h < a^m(\hat{\omega}, \hat{\eta})$. Notice that for $\eta \to 0$, $a_t^* \to \infty$, so $V'(a_t^*) = -D'(a_t^*)Q(a_t) - D(a_t^*)Q'(a_t^*) = -C'^{-1'}[Q(a_t^*)/\zeta]Q(a_t^*)Q'(a_t^*) - D(a_t^*)Q'(a_t^*) \to 0$. There exists an η' sufficiently small, such that:

$$V'(a_t^*) < \frac{1}{2}V'(\hat{a}^h) < \frac{1}{2}V(a^h) \text{ for all } a^h < \hat{a}^h$$

Therefore, for $\eta < \min\{\hat{\eta}, \eta'\}$,

$$\frac{\partial B(a^h)}{\partial a^h}|_{a^h \leq \hat{a}^h} < \omega \cdot \Big[V'(a_t^*) - V'(a^h) \Big]$$

$$< -\frac{1}{2}\omega V'(a^h).$$

Therefore,

$$B(a^h = 0) = B(\hat{a}^h) - \int_0^{\hat{a}^h} \frac{\partial B(x)}{\partial x} dx$$

$$> B(\hat{a}^h) + \frac{1}{2}\omega \int_0^{\hat{a}^h} V'(x)dx$$

$$> -\bar{\lambda}_H K + \frac{1}{2}\omega \int_0^{\hat{a}^h} V'(x)dx.$$

We can find an $\omega' > 0$ such that for $\omega > \omega'$, $\frac{1}{2}\omega \int_0^{\hat{a}^h} V'(x)dx > \mu_{t-1}\bar{\lambda}_H K$. Therefore, for $\eta < \min\{\hat{\eta}, \eta'\}$ and $\omega > \omega'$, $B(a^h = 0) > 0$.

The proof for the rest of the proposition also follows the same strategy as Proposition 1. $\square$

Proposition 6. *Suppose that a Mandarin starts with a sufficiently strong belief in Confucianism and a sufficiently weak belief in Taoism: $\mu_0^c > \frac{1}{2}$ and $\mu_0^d < \underline{\mu}$.*

1. *There exists an $I < \infty$, if the Mandarin is persecuted fewer than I times, the Mandarin's resistance a_t^* increases over each persecution. Therefore, the emperor's domination D_t^* decreases over each persecution.*

2. *After the $(I+1)$-th persecution, the Mandarin's belief in Confucianism is $\mu_{t-1}^c = 0$.*

3. *If the Mandarin has been persecuted more than $I+2$ times, the Mandarin's resistance a_t^**

decreases over each new persecution, driving the emperor's domination D_t^* to increase over each new persecution.

Proof. Denote $V(a_t) = -D(a_t)Q(a_t)$. For notational simplicity also ignore the subscript e in the secular cost function, so that $C_e(a_t) = C(a_t)$. Denote:

$$B(a^c, \mu) = \max_a \Big\{ \omega[V(a) - \eta C(a)] + H(a_t - a^c; \mu) \Big\} - \mu \bar{\lambda}_C K - \omega[V(a^c) - \eta C(a^c)].$$

Case 1: $B(a^c, \mu_0) < 0$.

At $\mu_t = \{\mu_0^c, \mu_0^d\}$, which includes $t = 1$, there are three candidates for the individual's optimal action. First, the individual can choose $a_0^* = \arg\max_a \omega[V(a) - \eta C(a)] + H(a_t - a^c; \mu_0^c)$, which yields a payoff of:

$$\omega[V(a_0^*) - \eta C(a_0^*)] + H(a_0^* - a^c; \mu_0^c) - [\mu_0^c \bar{\lambda}_C + \mu_0^d \bar{\lambda}_D + (1 - \mu_0^c - \mu_0^d)\lambda_M]K.$$

Second, the individual can choose a^c and obtains:

$$\omega[V(a^c) - \eta C(a^c)] - [\mu_t^d \bar{\lambda}_D + (1 - \mu_t^c - \mu_t^d)\lambda_M]K.$$

Third, the individual can choose a^d and obtains:

$$\omega[V(a^d) - \eta C(a^d)] + H(a^d - a^c; \mu_0^c) - [\mu_t^c \bar{\lambda}_C + (1 - \mu_t^c - \mu_t^d)\lambda_M]K.$$

Notice that first, the payoff of choosing a^c is higher than choosing a_0^*:

$$\omega[V(a_0^*) - \eta C(a_0^*)] + H(a_0^* - a^c; \mu_0^c) - [\mu_0^c \bar{\lambda}_C + \mu_0^d \bar{\lambda}_D + (1 - \mu_0^c - \mu_0^d)\lambda_M]K$$

$$\leq \omega[V(a^c) - \eta C(a^c)] - [\mu_0^d \bar{\lambda}_D + (1 - \mu_0^c - \mu_0^d)\lambda_M]K,$$

because $B^c(\mu_0) < 0$.

Second, the payoff of choosing a^c is higher than choosing a^d:

$$\omega[V(a^c) - \eta C(a^c)] - [\mu_t^d \bar{\lambda}_D + (1 - \mu_t^c - \mu_t^d)\lambda_M]K \geq \omega[V(a^d) - \eta C(a^d)] + H(a^d - a^c; \mu_0^c) - [\mu_t^c \bar{\lambda}_C + (1 - \mu_t^c - \mu_t^d)\lambda_M]K,$$

$$\omega[V(a^c) - \eta C(a^c)] - \omega[V(a^d) - \eta C(a^d)] - H(a^d - a^c; \mu_0^c) + [\mu_0^c \bar{\lambda}_C - \mu_0^d \bar{\lambda}_D]K > 0,$$

because $\omega[V(a^c) - \eta C(a^c)] > \omega[V(a^d) - \eta C(a^d)]$.

Therefore, the individual chooses

$$a_1^* = a^c.$$

Denote τ as the first period that the individual suffers from a disaster:

$$\mu_\tau^c = \frac{\mu_0^c \cdot 0}{\mu_0^c \cdot 0 + \mu_0^d \bar{\lambda}_D + (1 - \mu_0^c - \mu_0^d)\lambda_M} = 0,$$

$$\mu_\tau^d = \frac{\mu_0^d \bar{\lambda}_D}{\mu_0^d \bar{\lambda}_D + (1 - \mu_0^c - \mu_0^d)\lambda_M} > 0,$$

$$\mu_\tau^m = \frac{(1 - \mu_0^c - \mu_0^d)\lambda_M}{\mu_0^d \bar{\lambda}_D + (1 - \mu_0^c - \mu_0^d)\lambda_M} > 0.$$

The individual's optimization problem after $t \geq \tau + 1$ is the choice between

$$\omega[V(a_t^*) - \eta C(a_t^*)] + H(a_t^* - a^d; \mu_{t-1}^d) - [\mu_t^d \bar{\lambda}_D + (1 - \mu_t^d)\lambda_M]K,$$

where

$$a_t^* = \arg\max_a [\omega[V(a_t^*) - \eta C(a_t^*)] + H(a_t^* - a^d; \mu_{t-1}^d)]$$

and

$$\omega[V(a^d) - \eta C(a^d)] - (1 - \mu_t^d)\lambda_M K.$$

The individual chooses the above a_t^* if and only if:

$$\omega[V(a_t^*) - \eta C(a_t^*)] + H(a_t^* - a^d; \mu_{t-1}^d) - [\mu_t^d \bar{\lambda}_D + (1 - \mu_t^d)\lambda_M]K$$

$$\geq \omega[V(a^d) - \eta C(a^d)] - (1 - \mu_t^d)\lambda_M K,$$

$$\omega[V(a_t^*) - \eta C(a_t^*)] + H(a_t^* - a^d; \mu_{t-1}^d) - \omega[V(a^d) - \eta C(a^d)] \geq \mu_{t-1}^d \bar{\lambda}_D K,$$

which is satisfied because:

$$\omega[V(a_t^*) - \eta C(a_t^*)] + H(a_t^* - a^d; 1) - \omega[V(a^d) - \eta C(a^d)] \geq \bar{\lambda}_C K > \mu_{t-1}^d \bar{\lambda}_D K.$$

Notice that the cutoff is determined by:

$$\omega[V(a_t^*) - \eta C(a_t^*)] + H(a_t^* - a^d; \mu_{t-1}^d) - \omega[V(a^d) - \eta C(a^d)] = \mu_{t-1}^d \bar{\lambda}_D K.$$

The same as in the baseline model:

$$\frac{da^d}{d\mu_{t-1}^d} = \frac{[\bar{\lambda}_D K - \frac{\partial}{\partial \mu_{t-1}^d} H(a_t^* - a^d; \mu_{t-1}^d)]}{\omega\Big\{[V'(a_t^*) - V'(a^d)] + [\eta C'(a^d) - \eta C'(a_t^*)]\Big\}} < 0.$$

Therefore, because we have:

$$\omega[V(a_t^*) - \eta C(a_t^*)] + H(a_t^* - a^d; 1) - \omega[V(a^d) - \eta C(a^d)] \geq \bar{\lambda}_C K > \mu_{t-1}^d \bar{\lambda}_D K,$$

and $a^d < a^m$, the above inequality must be true. Because the individual never chooses a^d and $\bar{\lambda}_D > \lambda_M$, the Taoist belief $\mu_t^d \to 1$ after the individual suffers from a sufficiently large number of disasters.

Case 2: $B(a^c, \mu_0^c) > 0$ **and** $B(a^c, 1) < 0$.

Denote $a(i)$ as the action, $\{\mu^c(i),\ \mu^d(i)\}$ as the belief, of the Mandarin after he suffers from disasters for i times.

First, we have shown that the individual never chooses a^d under any belief $\{\mu^c(i), \mu^d(i)\}$, even under a full faith in Taoism.

Second, we claim that the individual will choose a^c after a finite number of disasters. By contradiction, suppose this is not the case. Also, note that the individual also never chooses a^d. So every time the individual suffers from a new disaster, we have:

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

or

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

because $\bar{\lambda}_C > \bar{\lambda}_D$, $\bar{\lambda}_C > \lambda_M$. As $i \to \infty$, $\mu^c(i) \to 1$, and

$$\omega[V(a_t^*) - \eta C(a_t^*)] + H(a_t^* - a^c; \mu_{t-1}^c) \to \omega[V(a^*) - \eta C(a^*)] + H(a^* - a^c; 1) < \bar{\lambda}_C K.$$

There exists an I sufficiently large, such that for $i \geq I$,

$$\omega[V(a_t^*) - \eta C(a_t^*)] + H(a_t^* - a^c; \mu^c(i)) < \mu^c(i)\bar{\lambda}_C K.$$

The claim is established.

Denote $\bar{I}$ as the number of disasters that the individual suffers so that the individual chooses a^c for the first time. We claim that for any $i \leq \bar{I}$, $\mu^d(i) > 0$ and $\mu^m(i) = 1 - \mu^d(i) - \mu^c(i) > 0$. At $t = 1$, $\mu^d(0) > 0$ and $\mu^m(0) > 0$ by assumption. Suppose that at a specific $j \leq \bar{I}$, $\mu^d(j) > 0$ and $\mu^m(j) > 0$. We have:

$$\mu^d(j+1) = \frac{\mu^d(j)\bar{\lambda}_D}{\mu^c(j)\bar{\lambda}_C + \mu^d(j)\bar{\lambda}_D + [1 - \mu^c(j) - \mu^d(j)]\lambda_M} > 0,$$

$$\mu^m(j+1) = \frac{\mu^m(j)\lambda_M}{\mu^c(j)\bar{\lambda}_C + \mu^d(j)\bar{\lambda}_D + [1-\mu^c(j)-\mu^d(j)]\lambda_M} > 0.$$

Because $I < \infty$, we can apply mathematical induction and conclude that for any $i \leq I$, $\mu^d(i) > 0$ and $\mu^m(i) = 1 - \mu^d(i) - \mu^c(i) > 0$. Therefore, we have:

$$\mu^c(\bar{I}+1) = 0,\ \mu^d(\bar{I}+1) > \mu^d(\bar{I}) > 0,\ \mu^m(\bar{I}+1) > 0.$$

So after suffering $\bar{I}+1$ disasters, the decision problem for the individual is straightforward. This drives $a^*(i)$ to be monotonically decreasing in i, which drives up the equilibrium domination at $D_t^*(a(i))$. □

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: $a^h \in [\underline{a}(\theta, \bar{\lambda}_H K), \bar{a}(\theta, \bar{\lambda}_H K)]$ 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 e_t^* is strictly positive after a sufficiently large number of suffering. The long-run effort strictly increases with a^h if $a^h < a^m(\theta)$ and strictly decreases with a^h if $a^h > a^m(\theta)$.*
2. *Suppose that a^h is a salvation religion: $a^h \in [0,\infty) \backslash [\underline{a}(\theta, \bar{\lambda}_H K), \bar{a}(\theta, \bar{\lambda}_H K)]$ as in Proposition 1. Then commoners always exert zero efforts in the contest: $e_t^* = 0$.*

Proof. Given the modified religious worldview 14, it is straightforward that the commoner's belief strictly increases after suffering from a new disaster:

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

Under the structure of the game, because commoners experience common shocks, we can ignore the superscript i in μ^i. The commoner i's problem in choosing the effort e_t^i is:

$$\max_{e_t^i} -\Gamma(e_t^i) + M(e_t^i; e_t^{-i}) \max_{a_t^i} \left\{ v(\theta)F(a_t^i) - C(a_t^i) + H(a_t^i - a^h; \mu_{t-1}) - \Big[\mu_{t-1}\mathbf{1}\{a_t^i \neq a^h\}\bar{\lambda}_H + (1-\mu_{t-1})\lambda_M\Big] K \right\}$$

$$+ [1 - M(e_t^i; e_t^{-i})] \max_{a_t^i} \left\{ v(\theta)F(a_t^i) - C(a_t^i) + H(a_t^i - a^h; \mu_{t-1}) - \Big[\mu_{t-1}\bar{\lambda}_H + (1-\mu_{t-1})\lambda_M\Big] K \right\}$$

We first look at the case of a world-affirming religion (Corollary 1). Suppose that the commoner becomes an elite in the period t by succeeding in the status competition, which happens with a probability of $M(e_t^i; e_t^{-i})$. In this case, Corollary 1 shows that an elite will choose $a_t^i = a^h$ when μ_{t-1} is sufficiently close to 1, which happens after a sufficiently large number of sufferings. The payoff to such an elite is:

$$[v(\theta)F(a^h) - C(a^h)] - (1 - \mu_{t-1})\lambda_M K.$$

Suppose that the commoner fails the status competition. In this case, the action a_t^i cannot directly brings "salvation." He therefore chooses receives the following payoff:

$$v(\theta)F(a_t^*) - C(a_t^*) + H(a_t^* - a^h; \mu_{t-1}) - \Big[\mu_{t-1}\bar{\lambda}_H + (1 - \mu_{t-1})\lambda_M\Big]K,$$

where $a_t^* = \arg\max_a [v(\theta)F(a_t^*) - C(a_t^*) + H(a_t^* - a^h; \mu_{t-1})]$. In the long run, the commoner's problem in choosing the effort e_t^i is:

$$\max_{e_t^i} -C(e_t^i) + M(e_t^i; e_t^{-i})\Bigg\{[v(\theta)F(a^h) - C(a^h)] - (1 - \mu_{t-1})\lambda_M K\Bigg\}$$

$$+[1 - M(e_t^i; e_t^{-i})]\Bigg\{v(\theta)F(a_t^*) - C(a_t^*) + H(a_t^* - a^h; \mu_{t-1}) - \Big[\mu_{t-1}\bar{\lambda}_H + (1 - \mu_{t-1})\lambda_M\Big]K\Bigg\}$$

$$= \max_{e_t^i} -\Gamma(e_t^i) - (1 - \mu_{t-1})\lambda_M K + M(e_t^i; e_t^{-i})[v(\theta)F(a^h) - C(a^h)] +$$

$$[1 - M(e_t^i; e_t^{-i})]\Big[v(\theta)F(a_t^*) - C(a_t^*) + H(a_t^* - a^h; \mu_{t-1}) - \mu_{t-1}\bar{\lambda}_H K\Big].$$

The payoff from winning the contest is strictly positive if and only if:

$$v(\theta)F(a^h) - C(a^h) > v(\theta)F(a_t^*) - C(a_t^*) + H(a_t^* - a^h; \mu_{t-1}) - \mu_{t-1}\bar{\lambda}_H K,$$

which is true because the religion is world affirming (Corollary 1).

In this case, the first order condition to e_t^i is:

$$\Gamma'(e_t^i) = \frac{\partial M}{\partial e_t^i} \underbrace{\Bigg\{[v(\theta)F(a^h) - C(a^h)] - [v(\theta)F(a_t^*) - C(a_t^*) + H(a_t^* - a^h; \mu_{t-1})] + \mu_{t-1}\bar{\lambda}_H K\Bigg\}}_{\Delta}.$$

The best response function of e_t^i to e_t^{-i} has the following slope:

$$\frac{de_t^i}{de_t^{-i}} = \frac{\overbrace{\frac{\partial^2 M}{\partial e_t^i \partial e_t^{-i}}}^{<0} \overbrace{\left\{[v(\theta)F(a^h) - C(a^h)] - [v(\theta)F(a_t^*) - C(a_t^*) + H(a_t^* - a^h; \mu_{t-1})] + \mu_{t-1}\bar{\lambda}_H K\right\}}^{>0}}{-\underbrace{\frac{\partial^2 M}{\partial (e_t^i)^2}}_{<0} \underbrace{\left\{[v(\theta)F(a^h) - C(a^h)] - [v(\theta)F(a_t^*) - C(a_t^*) + H(a_t^* - a^h; \mu_{t-1})] + \mu_{t-1}\bar{\lambda}_H K\right\}}_{>0} + \underbrace{\Gamma''(e_t^i)}_{>0}} < 0.$$

Therefore, there is a unique symmetric equilibrium where $e_t^i = e_t^{-i}$ for all i. Denote the equilibrium level of effort as e_t^*, which satisfies:

$$\Gamma'(e_t^*) = \frac{\partial M(e_t^*, e_t^*)}{\partial e_t^i} \left\{[v(\theta)F(a^h) - C(a^h)] - [v(\theta)F(a_t^*) - C(a_t^*) + H(a_t^* - a^h; \mu_{t-1})] + \mu_{t-1}\bar{\lambda}_H K\right\}. \tag{16}$$

The derivative of e_t^* with respect to a^h is derived as follows:

$$\Gamma'' de_t^* = \Big[\frac{\partial^2 M(e_t^*, e_t^*)}{\partial (e_t^i)^2} + \frac{\partial^2 M(e_t^*, e_t^*)}{\partial e_t^i \partial e_t^{-i}}\Big] \Delta de_t^* + \frac{\partial M(e_t^*, e_t^*)}{\partial e_t^i} \left\{[v(\theta)F'(a^h) - C'(a^h)] + H'(a_t^* - a^h; \mu_{t-1})\right\} da^h.$$

$$\left\{\Gamma'' - \Big[\frac{\partial^2 M(e_t^*, e_t^*)}{\partial (e_t^i)^2} + \frac{\partial^2 M(e_t^*, e_t^*)}{\partial e_t^i \partial e_t^{-i}}\Big] \Delta\right\} de_t^* = \frac{\partial M(e_t^*, e_t^*)}{\partial e_t^i} \left\{[v(\theta)F'(a^h) - C'(a^h)] + H'(a_t^* - a^h; \mu_{t-1})\right\} da^h,$$

$$\frac{de_t^*}{da^h} = \frac{\frac{\partial M(e_t^*, e_t^*)}{\partial e_t^i} \left\{[v(\theta)F'(a^h) - C'(a^h)] + H'(a_t^* - a^h; \mu_{t-1})\right\}}{\Gamma'' - \Big[\frac{\partial^2 M(e_t^*, e_t^*)}{\partial (e_t^i)^2} + \frac{\partial^2 M(e_t^*, e_t^*)}{\partial e_t^i \partial e_t^{-i}}\Big] \Delta}.$$

The first order condition of a_t^* is:

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

$$\frac{de_t^*}{da^h} = \frac{\frac{\partial M(e_t^*, e_t^*)}{\partial e_t^i} \left\{[v(\theta)F'(a^h) - C'(a^h)] - [v(\theta)F'(a^*) - C'(a^*)]\right\}}{\Gamma'' - \Big[\frac{\partial^2 M(e_t^*, e_t^*)}{\partial (e_t^i)^2} + \frac{\partial^2 M(e_t^*, e_t^*)}{\partial e_t^i \partial e_t^{-i}}\Big] \Delta}$$

$$= \frac{\overbrace{\frac{\partial M(e_t^*, e_t^*)}{\partial e_t^i}}^{>0} \Big\{ [v(\theta)F'(a^h) - v(\theta)F'(a^*) + C'(a^*) - C'(a^h)] \Big\}}{\underbrace{\Gamma'' - \Big[\frac{\partial^2 M(e_t^*, e_t^*)}{\partial (e_t^i)^2} + \frac{\partial^2 M(e_t^*, e_t^*)}{\partial e_t^i \partial e_t^{-i}}\Big]\Delta}_{>0}}.$$

For $a^h > a^m$, we have $a^h > a^*$, so $C'(a^*) - C'(a^h) < 0$ ($C'' > 0$) and $v(\theta)F'(a^h) - v(\theta)F'(a^*) \leq 0$ ($F''le0$). Therefore $\frac{de_t^*}{da^h} < 0$.

For $a^h < a^m$, we have $a^h < a^*$, so $C'(a^*) - C'(a^h) > 0$ and $v(\theta)F'(a^h) - v(\theta)F'(a^*) \geq 0$ ($F'' \geq 0$). Therefore $\frac{de_t^*}{da^h} > 0$.

Finally, we show that the equilibrium effort weakly increases with μ_{t-1}^i. Suppose that μ_{t-1}^i small enough so that an elite does not choose a^h. In this case, the elite chooses the same action as a commoner, so there is no benefit from the elite status. The effort level of all commoners is therefore zero. For μ_{t-1}^i large enough so that the elite chooses a^h, the equilibrium effort is identified by Condition 16. In this case,

$$\Gamma'(e_t^*) = \frac{\partial M(e_t^*, e_t^*)}{\partial e_t^i} \Big\{ [v(\theta)F(a^h) - C(a^h)] - [v(\theta)F(a_t^*) - C(a_t^*) + H(a_t^* - a^h; \mu_{t-1})] + \mu_{t-1}\bar{\lambda}_H K \Big\}.$$

$$\Gamma'' de_t^* = \Big[\frac{\partial^2 M(e_t^*, e_t^*)}{\partial (e_t^i)^2} + \frac{\partial^2 M(e_t^*, e_t^*)}{\partial e_t^i \partial e_t^{-i}}\Big]\Delta de_t^* + \frac{\partial M(e_t^*, e_t^*)}{\partial e_t^i}[\bar{\lambda}_H K - \frac{\partial H(a_t^* - a^h; \mu_{t-1})}{\partial \mu_{t-1}}] d\mu_{t-1}.$$

$$\frac{de_t^*}{d\mu_{t-1}} = \frac{\overbrace{\Gamma'' - \Big[\frac{\partial^2 M(e_t^*, e_t^*)}{\partial (e_t^i)^2} + \frac{\partial^2 M(e_t^*, e_t^*)}{\partial e_t^i \partial e_t^{-i}}\Big]\Delta}^{>0}}{\underbrace{\frac{\partial M(e_t^*, e_t^*)}{\partial e_t^i}}_{>0}[\bar{\lambda}_H K - \underbrace{\frac{\partial H(a_t^* - a^h; \mu_{t-1})}{\partial \mu_{t-1}}}_{<0}]} > 0.$$

The commoners' equilibrium effort level weakly increases with belief $\mu_{t-1} \in (0, 1)$.

Lastly, we look at the case of salvation religions such that $a^h \in [0, \infty] \backslash [\underline{a}(\theta, \bar{\lambda}_H K), \bar{a}(\theta, \bar{\lambda}_H K)]$. In this case, elites who just succeeded in the contest choose an action that cannot attain "salvation." Therefore, the payoff to a commoner at the contest stage is:

$$v(\theta)F(a_t^*) - C(a_t^*) + H(a_t^* - a^h; \mu_{t-1}) - \Big[\mu_{t-1}\bar{\lambda}_H + (1 - \mu_{t-1})\lambda_M\Big]K - \Gamma(e_t^i),$$

so commoners always chooses effort level at $e_t^* = 0$. □

C 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 "spiritual" power that enables religious figures to constrain political domination, especially by comparing the "spiritual powers" of the three major ideal-type religions.

C.1 How religious activism reacts to political persecution

C.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, the kings 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 were 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)).

C.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 was 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 *Rechtsstat*[...] (pp.41-42, Ullmann (1975)).

C.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.[11] A Chinese emperor can easily induce his Mandarins to "give up" by persecuting the Mandarins, who are bound to internally collapse and embrace Taoism.[12] 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. Finer actually ponders over 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

[11] These constraints are well articulated by, for example, Mengzi (2008); see also Lee (1992) and Tu (1993).

[12] 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 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)).

D Status competitions and world adjustment

World affirmation and status competitions in history In Imperial China, the Confucian institution of the civil service examination is 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 7 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 $i \in [0,1]$ choose his action a_t^i to maximize

$$v(\theta)F(a_t^i) - C(a_t^i) - [\mu_{t-1}\bar{\lambda}_H + (1-\mu_{t-1})\lambda_M]K.$$

The guilt-free commoner chooses $a^m(\theta)$, the optimal secular action. In turn, the commoner's effort e_t^i in the status competition maximizes the following function:

$$-\Gamma(e_t^i) + [1 - \Pi(e_t^i; e_t^{-i})]\Big\{v(\theta)F(a^m) - C(a^m) - [\mu_{t-1}\bar{\lambda}_H + (1-\mu_{t-1})\lambda_M]K\Big\} +$$

$$\Pi(e_t^i; e_t^{-i}) \max_a \Big\{B + v(\theta)F(a) - C(a) + H(a - a^h; \mu_{t-1}) - [\mu_{t-1}\bar{\lambda}_H \mathbf{1}\{a \neq a^h\} + (1-\mu_{t-1})\lambda_M]K\Big\}.$$

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. The next proposition formalizes the alignment between world adjustment and status competition under the modified setup.

Proposition 8. *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 $a^h \in (\underline{a}^s, \bar{a}^s)$, $de^*/da^h > 0$ for $a^h < a^m(\theta)$ and $de^*/da^h < 0$ for $a^h > a^m(\theta)$.

Proof. Define the (long-run) return to status as:

$$\Delta(a^h) = \begin{cases} B + v(\theta)F(a^h) - C(a^h) + \bar{\lambda}_H K - [v(\theta)F(a^m) - C(a^m)] & \text{for } a^h \in [\underline{a}(\theta, \bar{\lambda}_H K), \bar{a}(\theta, \bar{\lambda}_H K)] \\ B + v(\theta)F(a^*) - C(a^*) + H(a^* - a^h; 1) - [v(\theta)F(a^m) - C(a^m)] & \text{for } a^h \notin [\underline{a}(\theta, \bar{\lambda}_H K), \bar{a}(\theta, \bar{\lambda}_H K)] \end{cases}.$$

Notice that $\Delta(a^h)$ is continuous everywhere because $\lim_{a^h \to (\underline{a}(\theta, \bar{\lambda}_H K))^-} \Delta(a^h) = \Delta(\underline{a}(\theta, \bar{\lambda}_H K))$ and $\lim_{a^h \to (\bar{a}(\theta, \bar{\lambda}_H K))^+} \Delta(a^h) = \Delta(\bar{a}(\theta, \bar{\lambda}_H K))$.

Consider $a^h \in [\underline{a}(\theta, \bar{\lambda}_H K), \bar{a}(\theta, \bar{\lambda}_H K)]$, so that a^h is a religion of world adjustment. Notice that for $a^h \to a^m$, $\Delta(a^h) > 0$. In addition,

$$\Delta'(a^h) = v(\theta)F'(a^h) - C'(a^h) \gtreqless 0 \text{ for } a^h \lesseqgtr a^m.$$

Therefore, there exists a unique $A_1 < a^m$ and a unique $A_2 > a^m$, such that $\Delta(a^h) > 0$ if and only if $a^h \in (A_1, A_2)$.

Consider $a^h \notin [\underline{a}(\theta, \bar{\lambda}_H K), \bar{a}(\theta, \bar{\lambda}_H K)]$, so that a^h is a salvation religion. Notice that

$$\Delta'(a^h) = -H'(a^* - a^h; 1) \gtreqless 0 \text{ for } a^h \lesseqgtr a^m.$$

Therefore, there exists a unique $a_1 \leq \underline{a}(\theta, \bar{\lambda}_H K)$ and a unique $a_2 \geq \bar{a}(\theta, \bar{\lambda}_H K)$, such that $\Delta(a^h) < 0$ if and only if $a^h \in [0, a_1) \cup [a_2, \infty)$.

To summarize, there exists a unique $\underline{a}^s < a^m$ and a unique $\bar{a}^s > a^m$ such that $\Delta(a^h) > 0$ if and only if $a^h \in (\underline{a}^s, \bar{a}^s)$.

Now suppose $a^h \in (\underline{a}^s, \bar{a}^s)$. The first order condition for e_t^i is:

$$\frac{\partial}{\partial e_t^i}\Pi(e_t^i, e_t^{-i})\Delta(a^h) - \Gamma'(e_t^i) = 0.$$

The best response of e_t^i to e_t^{-i} has a negative slope:

$$\frac{de_t^i}{de_t^{-i}} = \frac{\frac{\partial}{\partial(e_t^i)\partial e_t^{-i}}\Pi(e_t^i, e_t^{-i})\Delta(a^h)}{\Gamma'(e_t^i) - \frac{\partial}{\partial(e_t^i)^2}\Pi(e_t^i, e_t^{-i})\Delta(a^h)} < 0.$$

Therefore, there exists a unique symmetric equilibrium, denoted as e_t^*, that satisfies:

$$\frac{\partial}{\partial e_t^i}\Pi(e_t^*, e_t^*)\Delta(a^h) - \Gamma'(e_t^*) = 0.$$

We can show that:

$$\frac{de_t^*}{da^h} = \frac{\frac{\partial}{\partial e_t^i}\Pi(e_t^*; e_t^*)\Delta'(a^h)}{\Gamma''(e_t^*) - [\frac{\partial}{\partial(e_t^i)^2}\Pi(e_t^*; e_t^*) + \frac{\partial}{\partial e_t^i \partial e_t^{-i}}\Pi(e_t^*; e_t^*)]} \lesseqgtr 0 \text{ for } \Delta'(a^h) \lesseqgtr 0.$$

Therefore, $de_t^*/da^h > 0$ for $a^h < a^m$ and $de_t^*/da^h < 0$ for $a^h > a^m$. □

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.

E Religious commandments on a forward-looking player

This section shows that a salvation religion has to impose a stricter commandment a^h on a forward-looking individual than a myopic individual. Therefore, religions of world adjustment remain unstable, and there can still be only two opposite types of salvation religions. The extra strictness is necessary because there are extra incentives for a forward-looking individual to obey the religious commandment.

Specifically, assume that the discount factor of the individual is $\delta \in [0, 1)$. For a forward-looking individual, we need to characterize the complete action plan under any belief μ. The complete plan can be identified by the value function in equilibrium:

$$V(\mu),$$

The value function $V(\mu)$ is the payoff to an individual at the start of a period when his belief that the religious worldview is correct is μ. We need to first show the existence and uniqueness of $V(\mu)$. This technical analysis then allows a substantive analysis. Specifically, I will identify the necessary and sufficient condition for a "never-falsifying equilibrium," defined as follows.

Definition 2. *An equilibrium is never falsifying if and only if the individual never chooses a^h on any realized equilibrium path.*

We compare the never-falsifying equilibrium for an individual with $\delta > 0$ and the never-falsifying equilibrium for an individual with $\delta = 0$. We will show that the distance between the religious commandment and the secular optimal action $|a^h - a^m|$ must be larger for a forward-looking individual than the myopic individual.

Proposition 9.

1. *For any $\delta < 1$, there is a unique equilibrium value function $V(\mu)$, corresponding to the unique optimal plan for the individual.*
2. *In a never-falsifying equilibrium, for the same belief, the religious commandment a^h is strictly more demanding for a forward-looking individual than for a myopic individual.*
3. *Denote B as the set of all religious commandments that supports the never-falsifying equilibrium. The set B for a forward-looking individual is a subset of the set B for a myopic individual.*
4. *If $a^h \in B$, a^h is a stable religion; if $a^h \notin B$, a^h is an unstable religion.*

Proof. **Ruling out suboptimal actions**

Denote:

$$a^*(\mu) = \arg\max_a [v(\theta)F(a) - C(a) + H(a - a^h; \mu)].$$

We first show that in any equilibrium, it is impossible for an individual to choose

$$a^{**} \neq a^*(\mu) \text{ and } a^* \neq a^h,$$

By contradiction, suppose that there is an equilibrium where at a belief μ:

$$a^{**} = a',\ a' \neq a^*(\mu) \text{ and } a' \neq a^h.$$

The payoff under the conjectured equilibrium is:

$$V(\mu) = v(\theta)F(a') - C(a') + H(a' - a^h; \mu) + [\mu\bar{\lambda}_H + (1-\mu)\lambda_M][-K + \delta V(\frac{\mu\bar{\lambda}_H}{\mu\bar{\lambda}_H + (1-\mu)\lambda_M})] +$$

$$\Big\{1 - [\mu\bar{\lambda}_H + (1-\mu)\lambda_M]\Big\} \cdot \delta V(\mu), \tag{17}$$

where $V(\mu)$ is the value function at μ.

Consider the single deviation by choosing

$$a'' = a^*(\mu),$$

for one period. The payoff under the single deviation is:

$$v(\theta)F(a^*(\mu)) - C(a^*(\mu)) + H(a^*(\mu) - a^h; \mu) + [\mu\bar{\lambda}_H + (1-\mu)\lambda_M][-K + \delta V(\frac{\mu\bar{\lambda}_H}{\mu\bar{\lambda}_H + (1-\mu)\lambda_M})] +$$

$$\Big\{1 - [\mu\bar{\lambda}_H + (1-\mu)\lambda_M]\Big\} \cdot \delta V(\mu).$$

By construction, we have:

$$v(\theta)F(a^*(\mu)) - C(a^*(\mu)) + H(a^*(\mu) - a^h; \mu) > v(\theta)F(a') - C(a') + H(a' - a^h; \mu),$$

the sign is strict because $v(\theta)F(a) - C(a) + H(a - a^h; \mu_{t-1})$ is strictly concave in a.

The rest of the terms in $V(\mu)'$ are identical to the rest of the terms in $V(\mu)$. Specifically, with probability $1 - [\mu\bar{\lambda}_H + (1-\mu)\lambda_M]$, the individual does not suffer from a disaster. In the next period $t+1$, the individual reverses back to the strategy under the conjectured equilibrium, specifically choosing a' for the period $t+1$, therefore obtaining $V(\mu)$. Therefore, it is a strictly more desirable single deviation to choose $a_t = a''$ for one period. We conclude that there is no equilibrium for an individual to choose $a_t^* \neq a^*(\mu)$ and $a_t^* \neq a^h$.

So the only possible non-falsifying equilibrium is for the individual to choose:

$$a^*(\mu) = \arg\max_a [v(\theta)F(a) - C(a) + H(a - a^h; \mu)]$$

under a belief μ.

Existence and uniqueness of the value function

In any equilibrium, the value function $V(\mu)$ is:

$$V(\mu) = \max \Bigg\{ v(\theta)F(a^h) - C(a^h) + (1-\mu)\lambda_M[-K + \delta V(0)] + [1 - (1-\mu)\lambda_M] \cdot \delta V(\mu),$$

$$v(\theta)F(a^*(\mu)) - C(a^*(\mu)) + H(a^*(\mu) - a^h; \mu) + [\mu\bar{\lambda}_H + (1-\mu)\lambda_M][-K + \delta V(\frac{\mu\bar{\lambda}_H}{\mu\bar{\lambda}_H + (1-\mu)\lambda_M})] +$$

$$\{1 - [\mu\bar{\lambda}_H + (1-\mu)\lambda_M]\} \cdot \delta V(\mu) \Bigg\}.$$

We will show that the value function $V(\mu)$ satisfies Blackwell (1965)'s sufficient condition for the contraction mapping theorem. To do so, first notice that $V(\mu)$ is a bounded function:

$$V(\mu) \in \Bigg[-\frac{1}{1-\delta}[v(\theta)F(a^h) - C(a^h) + H(a^m(\theta) - a^h; 1) + \bar{\lambda}_H K], 0 \Bigg],$$

where $a^m(\theta) = \arg\max_a [v(\theta)F(a) - C(a)]$. So we focus on the following functional space:

$$S \equiv \Big\{ w : [0,1] \to \mathbb{R}, w \text{ is bounded} \Big\}.$$

with the sup-norm metrics:

$$\rho(v, w) = \| w - y \| = \sup_{\mu \in [0,1]} |w(\mu) - y(\mu)|.$$

The normed functional space is a Banach space. We then define the functional transformation M in the space (S, ρ) by:

$$M(w)(\mu) = \max \Bigg\{ v(\theta)F(a^h) - C(a^h) + (1-\mu)\lambda_M[-K + \delta w(0)] + [1 - (1-\mu)\lambda_M] \cdot \delta w(\mu),$$

$$v(\theta)F(a^*(\mu))-C(a^*(\mu))+H(a^*(\mu)-a^h;\mu)+[\mu\bar{\lambda}_H+(1-\mu)\lambda_M][-K+\delta w(\frac{\mu\bar{\lambda}_H}{\mu\bar{\lambda}_H+(1-\mu)\lambda_M})]+$$

$$\left.\{1-[\mu\bar{\lambda}_H+(1-\mu)\lambda_M]\}\cdot\delta w(\mu)\right\}.$$

We verify the first condition in Blackwell (1965). Fix any $w(\cdot)\in S$ and $y(\cdot)\in S$ with $w(\cdot)<y(\cdot)$. There are two cases to discuss. The first case is when the first term of $M(w)(\mu)$ is larger than the second term. Therefore,

$$M(w)(\mu)=v(\theta)F(a^h)-C(a^h)+(1-\mu)\lambda_M[-K+\delta w(0)]+[1-(1-\mu)\lambda_M]\cdot\delta w(\mu)$$

$$<v(\theta)F(a^h)-C(a^h)+(1-\mu)\lambda_M[-K+\delta y(0)]+[1-(1-\mu)\lambda_M]\cdot\delta y(\mu)$$

$$\leq M(y)(\mu).$$

Similarly, in the second case, the second term of $M(w)(\mu)$ is larger than the first term. Therefore,

$$v(\theta)F(a^*(\mu))-C(a^*(\mu))+H(a^*(\mu)-a^h;\mu)+[\mu\bar{\lambda}_H+(1-\mu)\lambda_M][-K+\delta w(\frac{\mu\bar{\lambda}_H}{\mu\bar{\lambda}_H+(1-\mu)\lambda_M})]+$$

$$\left.\{1-[\mu\bar{\lambda}_H+(1-\mu)\lambda_M]\}\cdot\delta w(\mu)\right\}.$$

$$v(\theta)F(a^*(\mu))-C(a^*(\mu))+H(a^*(\mu)-a^h;\mu)+[\mu\bar{\lambda}_H+(1-\mu)\lambda_M][-K+\delta y(\frac{\mu\bar{\lambda}_H}{\mu\bar{\lambda}_H+(1-\mu)\lambda_M})]+$$

$$\left.\{1-[\mu\bar{\lambda}_H+(1-\mu)\lambda_M]\}\cdot\delta y(\mu)\right\}.$$

$$\leq M(y)(\mu).$$

Gathering both cases, for any $\mu\in[0,1]$, we have $M(w)(\mu)<M(y)(\mu)$. The first condition of Blackwell (1965) is verified.

We now verify the second condition in Blackwell (1965). Fix a $\Delta>0$. We have:

$$M(w+\Delta)(\mu)=\max\left\{v(\theta)F(a^h)-C(a^h)+(1-\mu)\lambda_M\{-K+\delta[w(0)+\Delta]\}+\right.$$

$$[1-(1-\mu)\lambda_M]\cdot\delta[w(\mu)+\Delta],$$

$$v(\theta)F(a^*(\mu))-C(a^*(\mu))+H(a^*(\mu)-a^h;\mu)+[\mu\bar{\lambda}_H+(1-\mu)\lambda_M]\{-K+\delta[w(\frac{\mu\bar{\lambda}_H}{\mu\bar{\lambda}_H+(1-\mu)\lambda_M})+\Delta]\}$$

$$
\left. +\{1-[\mu\bar{\lambda}_H+(1-\mu)\lambda_M]\}\cdot\delta[w(\mu)+\Delta]\right\}
$$

$$
= \max\left\{ v(\theta)F(a^h)-C(a^h)+(1-\mu)\lambda_M[-K+\delta w(0)]+\right.
$$

$$
[1-(1-\mu)\lambda_M]\cdot\delta w(\mu)+\delta\Delta,
$$

$$
v(\theta)F(a^*(\mu))-C(a^*(\mu))+H(a^*(\mu)-a^h;\mu)+[\mu\bar{\lambda}_H+(1-\mu)\lambda_M][-K+\delta w(\frac{\mu\bar{\lambda}_H}{\mu\bar{\lambda}_H+(1-\mu)\lambda_M})]
$$

$$
\left. +\{1-[\mu\bar{\lambda}_H+(1-\mu)\lambda_M]\}\cdot\delta w(\mu)+\delta\Delta\right\}
$$

$$
\leq M(w)(\mu)+\delta\Delta.
$$

So the second condition of Blackwell (1965) is satisfied. To conclude, $M(w)(\mu)$ is a contraction with modulus δ. There exists a unique fixed point $V(\cdot)$, such that:

$$
V(\mu)=G(V)(\mu).
$$

The value function represents the unique optimal plan for the individual.

The set of all commandments a^h that supports a never falsifying equilibrium To analyze the never-falsifying equilibrium, we define

$$
\mu(i)
$$

as in the text, i.e., the belief after the individual has suffered from disasters for i times in a never-falsifying equilibrium. The belief $\mu(i)$ is well defined recursively:

$$
\mu(i)=\frac{\mu(i-1)\bar{\lambda}_H}{\mu(i-1)\bar{\lambda}_H+[1-\mu(i-1)]\lambda_M}, \text{ and } \mu(0)=\mu_0>0.
$$

Abusing the notation slightly,

$$
a^*(i)\equiv a^*(\mu(i)), V(i)\equiv V(\mu(i))
$$

where

$$
a^*(i)=\arg\max_a[v(\theta)F(a)-C(a)+H(a-a^h;\mu(i))]
$$

and $V(i)$ is the value function if the individual's belief is $\mu(i)$. Because $\{\mu(i)\}_{i=0}^{\infty}$ is well defined, so $\{a^*(i)\}_{i=0}^{\infty}$ and $\{V(i)\}_{i=0}^{\infty}$ are also well defined.

Conjecture that the individual chooses $a^*(i)$ for all $i \in \{0, 1, 2, ...\}$. That is, the individual never exactly obeys the religious commandment a^h. Fix a specific i. For an individual who has suffered from a disaster for i times, the value function $V(i)$ under the conjectured equilibrium is:

$$V(i) = v(\theta)F(a^*(i)) - C(a^*(i)) + H(a^*(i) - a^h; \mu) + [\mu(i)\bar{\lambda}_H + (1 - \mu(i))\lambda_M][-K + \delta V(i+1)] +$$

$$\{1 - [\mu(i)\bar{\lambda}_H + (1 - \mu(i))\lambda_M]\} \cdot \delta V(i).$$

To understand the value function, with probability $[\mu(i)\bar{\lambda}_H + (1 - \mu(i)\lambda_M]$, the individual suffers from a disaster at a cost of $-K$, and enters a new state where the individual has suffered from disasters for $i + 1$ times, with the value function $V(i + 1)$; with probability $1 - [\mu(i)\bar{\lambda}_H + (1 - \mu(i)\lambda_M]$, the individual does not suffer from a disaster, and he receives the same value function for the next period at $V(i)$.

Consider the single deviation of choosing a^h. The payoff is:

$$[V(i)]' = v(\theta)F(a^h) - C(a^h) + (1 - \mu(i))\lambda_M \left[-K + \delta \frac{v(\theta)F(a^m(\theta)) - C(a^m(\theta)) - \lambda_M K}{1 - \delta}\right] + [1 - (1 - \mu(i))\lambda_M] \cdot \delta V(i).$$

The individual receives a secular cost $v(\theta)F(a^h) - C(a^h)$. The individual avoids the cost of guilt because he has chosen a^h. The individual now suffers from a disaster with a reduced probability $(1 - \mu(i))\lambda_M$, at a cost $-K$. But suffering from a disaster will fully falsify the religious promise. For the period $t + 1$, the individual's belief of the religious promise is 0, which implies that the guilt parameter ρ is also 0. The individual will choose $a_s^* = a^m(\theta)$ for all $s \geq t + 1$, which gives him a payoff of:

$$\frac{v(\theta)F(a^m(\theta)) - C(a^m(\theta)) - \lambda_M K}{1 - \delta},$$

With probability $[1 - (1 - \mu(i))\lambda_M]$, the individual does not suffer from a disaster. He reverses back to the strategy on the equilibrium path, receiving $V(i)$ for the next period.

A necessary condition to support the never-falsifying equilibrium is:

$$V(i) \geq [V(i)]'.$$

$$v(\theta)F(a^*(i)) - C(a^*(i)) + H(a^*(i) - a^h; \mu) + [\mu(i)\bar{\lambda}_H + (1 - \mu(i))\lambda_M][-K + \delta V(i + 1)] +$$

$$\{1 - [\mu(i)\bar{\lambda}_H + (1-\mu(i))\lambda_M]\} \cdot \delta V(i) \geq$$

$$v(\theta)F(a^h) - C(a^h) + (1-\mu(i))\lambda_M \left[-K + \delta \frac{v(\theta)F(a^m(\theta)) - C(a^m(\theta)) - \lambda_M K}{1-\delta}\right] + [1-(1-\mu(i))\lambda_M] \cdot \delta V(i).$$

Some algebra shows that the necessary condition is equal to:

$$v(\theta)F(a^*(i)) - C(a^*(i)) + H(a^*(i) - a^h; \mu(i)) - [v(\theta)F(a^h) - C(a^h)]$$
$$\geq \mu(i)\bar{\lambda}_H \cdot K + \mu(i)\bar{\lambda}_H \delta[V(i) - V(i+1)] + (1-\mu(i))\lambda_M \delta[\frac{v(\theta)F(a^m(\theta)) - C(a^m(\theta)) - \lambda_M K}{1-\delta} - V(i+1)]. \tag{18}$$

Compare the necessary condition with the same condition for a myopic individual with $\delta = 0$:

$$v(\theta)F(a^*(i)) - C(a^*(i)) + H(a^*(i) - a^h; \mu(i)) - [v(\theta)F(a^h) - C(a^h)] \geq \mu(i)\bar{\lambda}_H \cdot K. \tag{19}$$

It is easy to verify that:

$$V(i) - V(i+1) \geq 0,$$

and

$$\frac{v(\theta)F(a^m(\theta)) - C(a^m(\theta)) - \lambda_M K}{1-\delta} - V(i+1) > 0.$$

So the condition 18 is more restrictive than the condition 19. To see that $V(i) \geq V(i+1)$, notice that we have $\mu(i) \geq \mu(i+1)$. The inequalities are strict for $\mu(i) < 1$. To see that $\frac{v(\theta)F(a^m(\theta)) - C(a^m(\theta)) - \lambda_M K}{1-\delta} > V(i+1)$, notice that $H < 0$ for all $\mu > 0$, $\bar{\lambda}_H > \lambda_M$, and $v(\theta)F(a^m(\theta)) - C(a^m(\theta)) > v(\theta)F(a) - C(a)$ for all $a \neq a^m(\theta)$. So an individual under a belief $\mu(i+1)$ receives a lower utility from guilt, deviation from secular optimal action, and also a higher belief in the probability of suffering.

Fix all parameters other than a^h. The full set of religious commandment a^h that supports a never-falsifying equilibrium is:

$$B \equiv \bigcap_{i=0}^{\infty} \Bigg\{ a \in \mathbb{R} : v(\theta)F(a^*(i)) - C(a^*(i)) + H(a^*(i) - a; \mu(i)) - [v(\theta)F(a) - C(a)]$$
$$\geq \mu(i)\bar{\lambda}_H K + (1-\mu(i))\lambda_M \delta[\frac{v(\theta)F(a^m(\theta)) - C(a^m(\theta)) - \lambda_M K}{1-\delta} - V(i+1)] + \mu(i)\bar{\lambda}_H \delta[V(i) - V(i+1)] \Bigg\}.$$

Suppose $a^h \in B$. By construction, the individual never chooses a^h. Therefore $\lim_{i\to\infty} \mu(i) = 1$.

Suppose $a^h \notin B$. By construction, there exists a $j \in \{0, 1, ...\}$ such that:

$$v(\theta)F(a^*(j)) - C(a^*(j)) + H(a^*(j) - a^h; \mu(j)) - [v(\theta)F(a^h) - C(a^h)]$$
$$< \mu(j)\bar{\lambda}_H K + (1-\mu(j))\lambda_M \delta[\frac{v(\theta)F(a^m(\theta)) - C(a^m(\theta)) - \lambda_M K}{1-\delta} - V(j+1)] + \mu(j)\bar{\lambda}_H \delta[V(j) - V(j+1)] \tag{20}$$

Choose j' as the minimal j such that the condition 20 is satisfied. The number j' is well defined. If $j' = 0$, the individual chooses a^h at μ_0, so the next suffering falsifies the religious promise.

If $j' \geq 1$, the individual does not choose a^h until the individual's belief rises to $\mu(j')$ after he has suffered from disasters for j' times. The individual chooses a^h at $\mu(j')$, so the next suffering falsifies the religious promise.

To conclude, for $a^h \notin B$,

$$\lim_{i\to\infty} \mu(i) = 0.$$

Finally, we can show that

$$B \subseteq \Big\{a \in R : v(\theta)F(a^*) - C(a^*) + H(a^* - a; 1) - [v(\theta)F(a) - C(a)] \geq \bar{\lambda}_H K\Big\}.$$

where

$$a^* = \arg\max_a [v(\theta)F(a) - C(a) + H(a - a^h; 1)]$$

So a non-falsifying equilibrium must impose a more restrictive religious commandment on a forward-looking individual than a myopic one.

To see this, for an a^h such that $a^h \in B$, a necessary condition is:

$$v(\theta)F(a^*(i)) - C(a^*(i)) + H(a^*(i) - a^h; \mu(i)) - [v(\theta)F(a^h) - C(a^h)] \geq$$

$$\mu(i)\bar{\lambda}_H \cdot K + \mu(i)\bar{\lambda}_H \delta[V(i) - V(i+1)] + (1-\mu(i))\lambda_M \delta[\frac{v(\theta)F(a^m(\theta)) - C(a^m(\theta)) - \lambda_M K}{1-\delta} - V(i+1)].$$

For $i \to \infty$, we have:

$$v(\theta)F(a^*) - C(a^*) + H(a^* - a^h; 1) - [v(\theta)F(a^h) - C(a^h)] \geq \bar{\lambda}_H \cdot K,$$

because $\lim_{i\to\infty} V(i) - V(i+1) = 0$ and $\frac{v(\theta)F(a^m(\theta)) - C(a^m(\theta)) - \lambda_M K}{1-\delta} - V(i+1)$ is bounded while $\lim_{i\to\infty}(1-\mu(i))\lambda_M\delta = 0$. The claim is established. $\square$

F A Bayesian forward-looking player

This appendix extends the baseline model in Section 2 to the most general case. We look at an individual who is forward-looking with the same stage payoff function as in the baseline model. Moreover, the individual updates his belief through the Bayes Rule whenever possible:

- For each period t, regardless of whether he suffers, the individual uses the Bayes Rule to update his belief μ_t on whether the religious worldview is correct.

We prove a result that is analogous to the main propositions of the baseline model (Proposition 1 and Proposition 2). To retain its vitality in the long run, a religion must impose a commandment a^h that is sufficiently far away from $a^m(\theta)$.

We adopt a similar definition of stable versus unstable religions.

Definition 3. *1. A religion a^h is Bayes-unstable if there exists a finite sequence of events, after the sequence the belief on the religion is zero.*

2. A religion a^h is Bayes-stable if the religion is not Bayes-unstable. That is, one cannot find a finite sequence of events, after which the belief on the religion is zero.

Intuitively, there is a strictly positive probability that a Bayes-unstable religion will eventually be abandoned by an individual. But a Bayes-stable religion will never be fully abandoned, therefore always retaining its potential to revitalize faith among its followers. We are ready to state the main result for this appendix, Proposition 10. The proof shows that Proposition 10 is driven by the assumption that $\bar{\lambda}_H < 1$, so a disobeying individual will not always be punished by his God. Proposition 10 therefore highlights the importance of a merciful god for a religion to retain Bayesian followers in "good times."

Even though the punishment of a disobeying individual is limited by mercy, the bliss of an obeying individual is complete ($\lambda_H = 0$ for $a_t = a^h$) because salvation religions must claim that they have diagnosed the ultimate source of suffering.

Proposition 10. *1. For any $\delta < 1$, there is a unique equilibrium value function $V(\mu)$, corresponding to the unique optimal plan for the individual.*

2. In a never-falsifying equilibrium, for the same belief, the religious commandment a^h is strictly more demanding for a forward-looking individual than for a myopic individual.

3. Denote B as the set of all religious commandments that supports the never-falsifying equilibrium. The set B for a forward-looking individual is a subset of the set B for a myopic individual.

4. If $a^h \in B$, a^h is a Bayes-stable religion; if $a^h \notin B$, a^h is a Bayes-unstable religion.

Proof. **Rule out sub-optimal actions**

We first show that in any equilibrium, it is impossible for an individual to choose

$$a_t^* \neq \arg\max_a [v(\theta)F(a) - C(a)] + H(a - a^h; \mu) \text{ and } a_t^* \neq a^h,$$

under any belief $\mu \in [0,1]$. By contradiction, conjecture that there is an equilibrium, such that under a belief μ,

$$a_t^* \neq \arg\max_a [v(\theta)F(a) - C(a)] + H(a - a^h; \mu) \text{ and } a_t^* \neq a^h.$$

The payoff under the conjectured equilibrium is:

$$V(\mu) = [v(\theta)F(a_t^*) - C(a_t^*)] + H(a_t^* - a^h; \mu) + [\mu\bar{\lambda}_H + (1-\mu)\lambda_M][-K + \delta V(\frac{\mu\bar{\lambda}_H}{\mu\bar{\lambda}_H + (1-\mu)\lambda_M})]$$
$$+\Big\{1 - [\mu\bar{\lambda}_H + (1-\mu)\lambda_M]\Big\} \cdot \delta V(\frac{\mu(1-\bar{\lambda}_H)}{\mu(1-\bar{\lambda}_H) + (1-\mu)(1-\lambda_M)}). \tag{21}$$

Consider the single deviation to $a' = \arg\max_a [v(\theta)F(a) - C(a)] + H(a - a^h; \mu)$, while keeping future strategies the same. Under the single deviation, the payoff is:

$$V(\mu) = v(\theta)F(a') - C(a')] + H(a' - a^h; \mu) + [\mu\bar{\lambda}_H + (1-\mu)\lambda_M][-K + \delta V(\frac{\mu\bar{\lambda}_H}{\mu\bar{\lambda}_H + (1-\mu)\lambda_M})]$$

$$+\Big\{1 - [\mu\bar{\lambda}_H + (1-\mu)\lambda_M]\Big\} \cdot \delta V(\frac{\mu(1-\bar{\lambda}_H)}{\mu(1-\bar{\lambda}_H) + (1-\mu)(1-\lambda_M)}).$$

By the strict concavity of the objective function $[v(\theta)F(a) - C(a)] + H(a - a^h; \mu)$ in a,

$$v(\theta)F(a') - C(a')] + H(a' - a^h; \mu)$$

$$= \max_a [v(\theta)F(a) - C(a)] + H(a - a^h; \mu)$$

$$> [v(\theta)F(a_t^*) - C(a_t^*)] + H(a_t^* - a^h; \mu).$$

The single deviation yields a strictly higher payoff. The conjectured strategy a_t^* cannot be an equilibrium strategy. Define:

$$a^*(\mu; \omega, a^m, a^h) = \arg\max_a [v(\theta)F(a) - C(a)] + H(a - a^h; \mu),$$

and

$$a^m = \arg\max[v(\theta)F(a) - C(a)].$$

it is easy to how that $a^* \in (a^m, a^h)$ if $a^m < a^h$ and $a^* \in (a^h, a^m)$ if $a^h < a^m$.

Existence and uniqueness of the equilibrium In any equilibrium, the value function $V(\mu)$, which is a function of the belief $\mu \in [0,1]$, is:

$$V(\mu) = \max\Bigg\{v(\theta)F(a^h) - C(a^h) + (1-\mu)\lambda_M[-K + \delta V(0)] + [1-(1-\mu)\lambda_M]\cdot\delta V(\frac{\mu(1-\bar{\lambda}_H)}{\mu(1-\bar{\lambda}_H) + (1-\mu)(1-\lambda_M)}),$$

$$[v(\theta)F(a^*) - C(a^*)] + H(a^* - a^h;\mu) + [\mu\bar{\lambda}_H + (1-\mu)\lambda_M][-K + \delta V(\frac{\mu\bar{\lambda}_H}{\mu\bar{\lambda}_H + (1-\mu)\lambda_M})]$$

$$+\{1 - [\mu\bar{\lambda}_H + (1-\mu)\lambda_M]\}\cdot\delta V(\frac{\mu(1-\bar{\lambda}_H)}{\mu(1-\bar{\lambda}_H) + (1-\mu)(1-\lambda_M)})\Bigg\}.$$

We will show that the value function $V(\mu)$ satisfies Blackwell (1965)'s sufficient condition for the contraction mapping theorem. To do so, first notice that $V(\mu)$ is a bounded function:

$$V(\mu) \in \Bigg[\frac{1}{1-\delta}[v(\theta)F(a^h) - C(a^h) + H(a^m(\theta) - a^h;\mu) - \bar{\lambda}_H K], \frac{1}{1-\delta}[v(\theta)F(a^m(\theta)) - C(a^m(\theta))]\Bigg].$$

where $a^m(\theta) = \arg\max_a[v(\theta)F(a) - C(a)]$. So we focus on the following functional space:

$$S \equiv \Big\{w : [0,1] \to \mathbb{R}, v \text{ is bounded}\Big\}.$$

with the sup-norm metrics:

$$\rho(w, y) = \| w - y \| = \sup_{\mu\in[0,1]} |w(\mu) - y(\mu)|.$$

The functional space of bounded functions with sup-norm metrics is a Banach space. We then define the functional transformation M in the space (S, ρ) by:

$$M(w)(\mu) = \max\Bigg\{v(\theta)F(a^h) - C(a^h) + (1-\mu)\lambda_M[-K + \delta w(0)] + [1-(1-\mu)\lambda_M]\cdot\delta w(\frac{\mu(1-\bar{\lambda}_H)}{\mu(1-\bar{\lambda}_H) + (1-\mu)(1-\lambda_M)}),$$

$$[v(\theta)F(a^*) - C(a^*)] + H(a^* - a^h;\mu) + [\mu\bar{\lambda}_H + (1-\mu)\lambda_M][-K + \delta w(\frac{\mu\bar{\lambda}_H}{\mu\bar{\lambda}_H + (1-\mu)\lambda_M})]$$

$$+\{1-[\mu\bar{\lambda}_H+(1-\mu)\lambda_M]\}\cdot\delta w(\frac{\mu(1-\bar{\lambda}_H)}{\mu(1-\bar{\lambda}_H)+(1-\mu)(1-\lambda_M)})\Bigg\}.$$

We verify the first condition in Blackwell (1965). Fix any $w(\cdot)\in S$ and $y(\cdot)\in S$ with $v(\cdot)<w(\cdot)$. There are two cases to discuss. The first case is when the first term of $M(v)(\mu)$ is larger than the second term. Therefore,

$$M(w)(\mu)=v(\theta)F(a^h)-C(a^h)+(1-\mu)\lambda_M[-K+\delta w(0)]+[1-(1-\mu)\lambda_M]\cdot\delta w(\frac{\mu(1-\bar{\lambda}_H)}{\mu(1-\bar{\lambda}_H)+(1-\mu)(1-\lambda_M)})$$

$$<v(\theta)F(a^h)-C(a^h)+(1-\mu)\lambda_M[-K+\delta y(0)]+[1-(1-\mu)\lambda_M]\cdot\delta y(\frac{\mu(1-\bar{\lambda}_H)}{\mu(1-\bar{\lambda}_H)+(1-\mu)(1-\lambda_M)})$$

$$\leq M(y)(\mu).$$

Similarly, in the second case, the second term of $M(w)(\mu)$ is larger than the first term. Therefore,

$$[v(\theta)F(a^*)-C(a^*)]+H(a^*-a^h;\mu)+[\mu\bar{\lambda}_H+(1-\mu)\lambda_M][-K+\delta w(\frac{\mu\bar{\lambda}_H}{\mu\bar{\lambda}_H+(1-\mu)\lambda_M})]$$

$$+\{1-[\mu\bar{\lambda}_H+(1-\mu)\lambda_M]\}\cdot\delta w(\frac{\mu(1-\bar{\lambda}_H)}{\mu(1-\bar{\lambda}_H)+(1-\mu)(1-\lambda_M)})$$

$$<[v(\theta)F(a^*)-C(a^*)]+H(a^*-a^h;\mu)+[\mu\bar{\lambda}_H+(1-\mu)\lambda_M][-K+\delta y(\frac{\mu\bar{\lambda}_H}{\mu\bar{\lambda}_H+(1-\mu)\lambda_M})]$$

$$+\{1-[\mu\bar{\lambda}_H+(1-\mu)\lambda_M]\}\cdot\delta y(\frac{\mu(1-\bar{\lambda}_H)}{\mu(1-\bar{\lambda}_H)+(1-\mu)(1-\lambda_M)})$$

$$\leq M(y)(\mu).$$

Gathering both cases, for any $\mu\in[0,1]$, we have $M(v)(\mu)<M(w)(\mu)$. The first condition of Blackwell (1965) is verified.

We now verify the second condition in Blackwell (1965). Fix a $\Delta>0$. We have:

$$M(w+\Delta)(\mu)=\max\Bigg\{v(\theta)F(a^h)-C(a^h)+(1-\mu)\lambda_M\{-K+\delta[w(0)+\Delta]\}+$$

$$[1-(1-\mu)\lambda_M]\cdot\delta[w(\frac{\mu(1-\bar{\lambda}_H)}{\mu(1-\bar{\lambda}_H)+(1-\mu)(1-\lambda_M)})+\Delta],$$

$$[v(\theta)F(a^*)-C(a^*)]+H(a^*-a^h;\mu)+[\mu\bar{\lambda}_H+(1-\mu)\lambda_M]\{-K+\delta[w(\frac{\mu\bar{\lambda}_H}{\mu\bar{\lambda}_H+(1-\mu)\lambda_M})+\Delta]\}$$

$$+\{1-[\mu\bar{\lambda}_H+(1-\mu)\lambda_M]\}\cdot\delta[w(\frac{\mu(1-\bar{\lambda}_H)}{\mu(1-\bar{\lambda}_H)+(1-\mu)(1-\lambda_M)})+\Delta]\Bigg\}$$

$$\leq M(w)(\mu)+\delta\Delta.$$

So the second condition of Blackwell (1965) is satisfied. To conclude, $M(w)(\mu)$ is a contraction with modulus δ. There exists a unique fixed point $V(\cdot)$, such that:

$$V(\mu)=M(V)(\mu).$$

The value function represents the unique equilibrium.

Construct the set of a^h that induces a never-falsifying equilibrium Suppose that the individual has experienced suffering for $i\geq 0$ times out of $i+j\geq 0$ times, and the individual has never chosen a^h. The probability of this event under the religious worldview is:

$$\bar{\lambda}_H^i(1-\bar{\lambda}_H)^j,$$

notice that the probability is independent of the sequence that the i sufferings arrives. Similarly, the probability of this event under the naturalist worldview is:

$$\lambda_M^i(1-\lambda_M)^i.$$

Define $\mu(i,j)$ as following, which is the posterior belief on the religious worldview if the individual has suffered for $i\geq 0$ out of $i+j$ times and he has never chosen a^h:

$$\mu(i,j)=\frac{\mu_0\bar{\lambda}_H^i(1-\bar{\lambda}_H)^j}{\mu_0\bar{\lambda}_H^i(1-\bar{\lambda}_H)^j+(1-\mu_0)\lambda_M^i(1-\lambda_M)^j}.$$

For an individual to not choose a^h under $\mu(i,j)$ in a never falsifying equilibrium, we need:

$$\overbrace{[v(\theta)F(a^*)-C(a^*)]}^{\text{secular payoff at } a_t^*=a^*}+\overbrace{H(a^*-a^h;\mu(i,j))}^{\text{guilt at } a_t^*=a^*}+\overbrace{[\mu(i,j)\bar{\lambda}_H+(1-\mu(i,j))\lambda_M]}^{\text{prob. suffering under } a_t^*=a^*\neq a^h}[-K+\delta V(\mu(i+1,j))]$$

$$+\overbrace{\{1-[\mu(i,j)\bar{\lambda}_H+(1-\mu(i,j))\lambda_M]\}}^{\text{prob. no suffering under } a_t^*=a^*\neq a^h}\cdot\delta V(\mu(i,j+1))\geq$$

$$\overbrace{v(\theta)F(a^h)-C(a^h)}^{\text{secular payoff at } a_t^*=a^h}+\overbrace{(1-\mu(i,j))\lambda_M}^{\text{prob. suffering under } a_t^*=a^h}[-K+\delta V(0)]+\overbrace{[1-(1-\mu(i,j))\lambda_M]}^{\text{prob. no suffering at } a_t^*=a^h}\cdot\delta V(\tilde{\mu}(i,j+1)),$$

where

$$\tilde{\mu}(i, j+1) = \frac{\mu_0 \bar{\lambda}_H^i (1-\bar{\lambda}_H)^j}{\mu_0 \bar{\lambda}_H^i (1-\bar{\lambda}_H)^j + (1-\mu_0)\lambda_M^i (1-\lambda_M)^{j+1}},$$

and

$$a^* \equiv a^*(\mu(i,j), a^h) = \arg\max_{\tilde{a}} \Big[v(\theta)F(\tilde{a}) - C(\tilde{a}) + H(\tilde{a} - a^h; \mu(i,j)) \Big].$$

or equivalently,

$$[v(\theta)F(a^*) - C(a^*)] + H(a^* - a^h; \mu(i,j)) - [v(\theta)F(a^h) - C(a^h)] \geq$$

$$(1-\mu(i,j))\lambda_M[-K+\delta V(0)] - [(1-\mu(i,j))\lambda_M][-K+\delta V(\mu(i+1,j))] - \mu(i,j)\bar{\lambda}_H[-K+\delta V(\mu(i+1,j))]$$

$$[1-(1-\mu(i,j))\lambda_M]\delta V(\tilde{\mu}(i,j+1)) - \{1-(1-\mu(i,j))\lambda_M\}\delta V(\mu(i,j+1)) + \delta\mu(i,j)\bar{\lambda}_H V(\mu(i,j+1))$$

$$= (1-\mu(i,j))\lambda_M[\delta V(0) - \delta V(\mu(i+1,j))] + \mu(i,j)\bar{\lambda}_H K - \mu(i,j)\bar{\lambda}_H \delta V(\mu(i+1,j))$$

$$+\delta\mu(i,j)\bar{\lambda}_H V(\mu(i,j+1)) + [1-(1-\mu(i,j))\lambda_M]\delta[V(\tilde{\mu}(i,j+1)) - V(\mu(i,j+1))]$$

$$= \mu(i,j)\bar{\lambda}_H K + (1-\mu(i,j))\lambda_M \delta[V(0) - V(\mu(i+1,j))]$$

$$+\mu(i,j)\bar{\lambda}_H\delta[V(\mu(i,j+1)) - V(\mu(i+1,j))] + [1-(1-\mu(i,j))\lambda_M]\delta[V(\tilde{\mu}(i,j+1)) - V(\mu(i,j+1))].$$

Therefore, define:

$$B_{ij} = \Bigg\{ a \geq 0 : [v(\theta)F(a^*) - C(a^*)] + H(a^* - a; \mu(i,j)) - [v(\theta)F(a) - C(a)] \geq$$

$$\mu(i,j)\bar{\lambda}_H K + (1-\mu(i,j))\lambda_M \delta[V(0) - V(\mu(i+1,j))]$$

$$+\mu(i,j)\bar{\lambda}_H\delta[V(\mu(i,j+1)) - V(\mu(i+1,j))] + [1-(1-\mu(i,j))\lambda_M]\delta[V(\tilde{\mu}(i,j+1)) - V(\mu(i,j+1))] \Bigg\}.$$

And then define:

$$B = \cap_{i=0}^{\infty} \cap_{j=0}^{\infty} B_{ij}.$$

One can see that a religious commandment a^h induces a non-falsifying equilibrium if and only if $a^h \in B$.

The equilibrium under $a^h \in B$ is Bayes-stable We have shown that for $a^h \in B$, the individual never chooses a^h. To show that for any $i, j \geq 0$, $\mu(i,j) > 0$, first notice that $\mu(i,j) \geq \mu(0,j)$. It is sufficient to show that $\mu(0,j) > 0$. For any fixed j:

$$\mu(0,j) = \frac{\mu_0(1-\bar{\lambda}_H)^j}{\mu_0(1-\bar{\lambda}_H)^j + (1-\mu_0)(1-\lambda_M)^j} > 0.$$

We conclude that for all $a^h \in B$, the religion commandment a^h is Bayes stable.

The equilibrium under $a^h \notin B$ is Bayes unstable Fix a commandment

$$a^h \notin B.$$

By construction, there exists a pair (i, j), such that:

$$a^h \notin B_{ij}.$$

If there are multiple such pairs, pick one as follows. First, pick all pairs such that $i + j$ is the minimal among all pairs. Second, if there are more than one pair with a minimal $i+j$, pick the pair with the minimal i. Denote such a pair as (i', j'). By construction, for all $i'' < i', j'' < j'$, we have $a^h \in B_{i''j''}$.

Consider the following sequence of events: the individual first experiences j' periods of no suffering, and then i' periods of suffering. The individual does not choose a^h for the first $i' + j' - 1$ periods, and chooses a^h for the period $i' + j'$. If the individual then suffers from a disaster in the period $i' + j'$, the individual's belief on the religious worldview is zero. We conclude that for $a^h \notin B$, a religion that sanctifies a^h is Bayes-unstable.

Show that the set B is smaller for a forward looking player than a myopic player ($\delta = 0$) To ensure that the individual does not choose a^h, a necessary condition is:

$$a^h \in B_{ij} \text{ for } i, j \geq 0.$$

Fix j, as $i \to \infty$,we have $\mu(i, j) \to 1$. Therefore,

$$[v(\theta)F(a^*) - C(a^*)] + H(a^* - a^h; \mu(i, j)) - [v(\theta)F(a^h) - C(a^h)] \geq$$

$$\mu(i,j)\bar{\lambda}_H K + \overbrace{(1-\mu(i,j))}^{\to 0}\lambda_M \delta[V(0) - V(\mu(i+1,j))]$$
$$+\mu(i,j)\bar{\lambda}_H \delta \overbrace{[V(\mu(i,j+1)) - V(\mu(i+1,j))]}^{\to 0 \text{ because } \mu(i,j+1)\to 1 \text{ and } \mu(i+1,j) \to 1}$$
$$+[1-(1-\mu(i,j))\lambda_M]\delta \overbrace{[V(\tilde{\mu}(i,j+1)) - V(\mu(i,j+1))]}^{\to 0 \text{ because } \tilde{\mu}(i,j+1)\to 1 \text{ and } \mu(i,j+1)\to 1}$$

Therefore, for a fixed j, as $i \to \infty$,

$$[v(\theta)F(a^*) - C(a^*)] + H(a^* - a^h; 1) - [v(\theta)F(a^h) - C(a^h)] \geq \bar{\lambda}_H K,$$

which is the necessary and sufficient condition for a myopic player with $\delta = 0$ to never choose a^h. The claim is established. $\square$