

The Moral Credentials of Divine Authority

— 42

Phil Stilwell

Independent Scholar

ORCID: [0009-0009-3725-8682](https://orcid.org/0009-0009-3725-8682)

Revised August 2026

Central claim

Before a person rationally defers morally to a candidate God-concept, there must be a credentials case whose force is independent of the disputed authority claim, that discriminates trustworthy from untrustworthy claimants, and that remains open to correction. Ordinary legitimate authority gains standing through relevant competence and character; power, creatorhood, threats, rewards, self-attestation, disputed texts, and total inscrutability do not supply this three-part bridge. The result is a prior-screening requirement, not a disproof of theism or a claim of human infallibility: a genuinely good deity may survive scrutiny, but no claimant may forbid scrutiny before legitimacy is established.

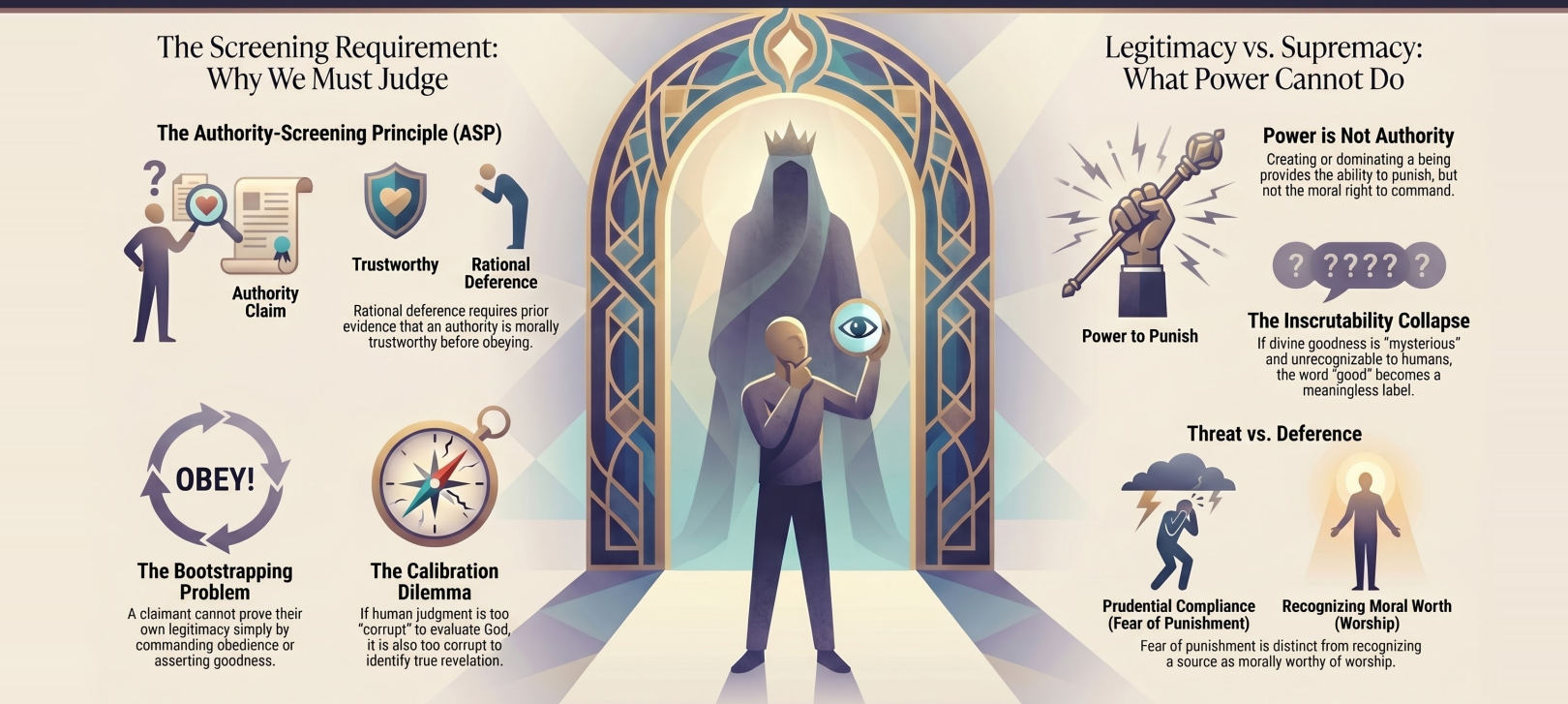
Preview:

Authority claim	Credentials test	Why shortcuts fail	Rational conclusion
A purported deity demands morally binding deference.	Evidence must be independent of the disputed claim, discriminating, and open to correction.	Status, force, self-certification, and opaque goodness either fail to distinguish character or disable correction.	Deference is warranted only after a candidate passes this fallible screening test.

The decisive question is not whether perfect goodness would merit trust, but what evidence responsibly identifies a purported authority as morally good before judgment is surrendered to it.

Visual preview:

The Gatekeeper of Faith: Why Moral Scrutiny Must Precede Divine Obedience



The preserved map's Authority-Screening Principle is specified here by the three-part moral-credentials test; its side panels distinguish bootstrapping, calibration, supremacy, inscrutability, and fear. It depicts the screening burden, not a verdict that every deity-claim fails.

Abstract

Rational moral deference to a candidate God-concept requires a prior, defeasible credentials case with three features: its force is independent of the disputed authority claim, it discriminates trustworthy from untrustworthy claimants, and it remains open to correction. This priority is structural, not a demand that every act of trust begin with certainty or religiously neutral evidence. People encounter purported ultimate authorities through rival traditions, texts, institutions, experiences, arguments, and alleged commands; they do not begin with an already certified perfectly good being. A cumulative comparison shows that power, creatorhood, reward, threat, self-attestation, and textual attribution can accompany both trustworthy and untrustworthy claimants, while total appeals to inscrutable goodness remove the contrast needed to identify moral goodness. Modified divine-command theories may explain how commands from a loving deity generate obligations, but they do not by themselves identify which candidate or interpretation expresses that character. Modest skepticism about human judgment counsels humility; global skepticism defeats warranted praise and revelation-identification as well as criticism. The result is a prior-screening requirement: evidence of truthful, just, compassionate, non-manipulative character must be independent of the disputed claim, discriminating, and publicly correctable before rational surrender of moral judgment. This conclusion does not disprove theism or make human judgment infallible. It states the burden that separates worship from capitulation to power.

Keywords: moral authority; divine command theory; candidate deity; moral credentials; obedience; epistemic authority; religious ethics; Euthyphro

1. The Decision That Comes Before Obedience

The claim that divine commands are morally binding is often presented from inside a settled tradition: the deity is perfectly good, the revelation is authentic, and obedience is therefore fitting. From within that package, distrust can look like rebellion. The more basic problem arises one step earlier. Finite persons encounter rival claims about ultimate authority through scriptures, prophets, institutions, inherited communities, miracles, philosophical arguments, mystical experiences, and private impressions. Before obedience is morally required, someone must decide which claimant, text, interpretation, or experience—if any—deserves trust.

Call a specified conception of a deity a *God-concept*. This term matters because the bare word normally gathers together very different claims about character, action, revelation, judgment, and human obligation. *Theism* is the general position that at least one deity exists. A God-concept adds content: what the deity is like, what it has done, which communications are attributed to it, and why its directives supposedly bind. The argument here concerns the rational evaluation of candidate God-concepts. It does not assume that every candidate describes the same being, that any candidate is real, or that every disagreement reduces to character.

The governing thesis is direct: rational moral deference to a candidate deity requires prior warrant that the candidate is morally trustworthy in the relevant domain. *Prior* names an evidential order, not necessarily a chronological one. The warrant need not be complete, religiously neutral, or indubitable, but its force cannot arise solely from the authority claim, directive, threat, or disputed revelation whose standing is under assessment. Trust can rationally outstrip understanding after a source becomes credible. Credibility cannot be manufactured by ordering the evaluator to stop evaluating.

Three conditions make that requirement concrete. Evidence must have *independence*: at least part of its force survives temporarily bracketing the claimant's demand and self-description, so the reasoning does not bootstrap from the disputed claim. It must have *discrimination*: the evidence is more expected if the claimant is trustworthy than if an untrustworthy rival merely asserts the same status. It must have *correctability*: adverse evidence can lower confidence in the candidate, channel, or interpretation. These conditions can be met cumulatively; no single observation must carry the whole case. Throughout the paper, a *moral-credentials case* means a body of evidence that passes this three-part test well enough for the stakes and domain of deference.

This question matters most when a purported directive bears on another person's welfare. To defer morally is not merely to accept help with a calculation. It is to let another's judgment guide one's conscience, loyalties,

and treatment of persons. Kant's account of practical agency resists reducing persons to instruments, while rights-based accounts of justice treat the worth of persons as imposing limits on what may be done to them (Kant, 1996; Wolterstorff, 2008). A mistaken forecast may waste time; a mistaken moral authority can sacralize cruelty. The required credentials should therefore rise with the moral stakes.

The intended reader may already believe that the correct deity is perfectly good. Nothing in the argument asks the reader to deny that belief merely because it is religious. It asks for one consistency: use enough moral and epistemic judgment to distinguish the true candidate from a manipulator, a mistaken institution, a cruel interpretation, or a human projection. If those distinctions cannot be made, responsible obedience cannot be identified either.

2. Supremacy, Standing, and Identification

Three questions are easily collapsed:

1. **Supremacy:** How powerful, knowledgeable, or causally basic is the claimant?
2. **Standing:** Does the claimant have the moral right to direct another person's agency in this domain?
3. **Identification:** What warrants believing that this encountered claimant, text, or interpretation has that standing?

The first does not automatically answer the second. The capacity to punish can make refusal costly without making punishment just. Causal authorship can generate responsibility for what is created without establishing ownership of persons. Superior knowledge can support epistemic trust in one domain while leaving moral character in another domain unsettled. A being able to dominate every outcome has more leverage, not thereby more legitimacy.

Contemporary work on authority explains why. Darwall describes moral obligation through a second-personal standpoint in which claims and demands presuppose standing between persons (Darwall, 2006). Enoch analyzes authority as a capacity to affect reasons, not merely behavior (Enoch, 2014). Zagzebski's defense of epistemic authority likewise does not say that confidence, rank, or power makes a source authoritative; deference is justified through reflective reasons for judging that following the source is more likely to lead one well in the relevant domain (Zagzebski, 2012). These accounts disagree in important ways, but all make authority normatively thicker than force.

The third question is then unavoidable. Suppose perfect goodness would confer moral standing on a deity. A finite evaluator still has to determine whether the candidate encountered through a particular channel possesses that goodness and whether the alleged directive has been correctly attributed. The perfect-being premise does not authenticate a manuscript, select among rival revelations, resolve translation, or rule out institutional self-interest. The label attached to the source cannot do the work of identifying the source.

The comparison with ordinary authority is useful but limited. A physician earns medical trust through training, records, peer accountability, and conduct; none makes error impossible. A judge has lawful standing only within a jurisdiction and remains constrained by evidence and procedure. A parent has special responsibilities toward a child but does not own the child. These cases are not blueprints for a deity. They show the familiar logic of responsible deference: relevant standing is supported by reasons that do not arise solely from the authority's demand to be trusted.

The Moral-Credentials Test

Rational moral deference to a purported authority requires sufficient, defeasible warrant that the authority is morally trustworthy in the domain of deference. A cumulative case must have (1) independence, so some evidential force survives bracketing the disputed authority claim; (2) discrimination, so it favors a trustworthy candidate over rivals able to mimic status or force; and (3) correctability, so contrary evidence can count against the candidate, channel, or interpretation. Credible character, truthfulness, proportionate judgment, non-manipulation, concern for persons, and reliable performance may contribute. The test permits deeper trust after credibility is established; it denies that a command can supply its own credentials.

This test leaves room for genuine authority. It opposes a lower evidential standard for ultimate authority than for ordinary authority precisely where the cost of error is greatest.

3. Why the Familiar Shortcuts Do Not Close the Gap

Several alleged credentials may be evidence of something important. The error is to treat what they show as identical to moral legitimacy. Table 1 shows where each shortcut fails the three-part test and what further bridge remains necessary.

Table 1: What alleged credentials can and cannot establish

Alleged credential	What it can show	Why it is not enough	What the moral gap still requires
Power or supremacy	Capacity, knowledge, causal reach, or ability to impose outcomes.	Force changes prudential options; it does not create moral standing.	Reasons connecting the claimant's character and purposes to justified authority.
Creatorhood	A special causal relation and perhaps duties of care or gratitude.	Causing a person to exist is not ownership of that person or unlimited permission.	A defensible account of the creator's rights, duties, and treatment of persons.
Self-attestation or command	What the claimant says about itself and what it directs.	The source's trustworthiness is the question; using its assertion alone is circular.	Evidence whose force does not depend on already granting the disputed authority.
Scriptural or institutional attribution	A tradition's testimony, continuity, and interpretive framework.	Rival texts and institutions make incompatible attributions and can certify themselves.	Source authentication, interpretation, and moral comparison that remain open to correction.
Threat or reward	The practical cost of refusal or benefit of compliance if the report is credible.	Incentive can explain compliance without justifying the threatener or promiser.	Proportionality, truthfulness, non-manipulation, and morally defensible ends.
Total inscrutability	A reminder that finite judgment can miss reasons and wider goods.	If every conduct profile fits perfect goodness equally, goodness no longer discriminates candidates.	Enough recognizable continuity with justice, love, and truth to support warranted praise.

The table does not show that every listed feature is irrelevant. It shows that none alone supplies a bridge with independence, discrimination, and correctability from status, testimony, or incentive to morally trustworthy standing.

3.1 Self-attestation and disputed texts

Self-attestation fails the independence condition when its force depends on already trusting the speaker. A claimant may say, in effect, "I am good; therefore obey." Testimony from a trusted source can provide evidence. Testimony from a source whose honesty is the question cannot settle that question by itself. The same applies

when a text says it comes from a perfectly good deity and its authority is then defended by citing the text's own claim. Other considerations may support the text, but the closed loop does not.

This is an ordinary anti-bootstrapping constraint, not hostility to revelation. A stranger does not become trustworthy merely by demanding trust. A government does not become legitimate merely by outlawing questions about legitimacy. A teacher does not establish competence by marking every challenge as disobedience. If a religious channel is reliable, it should be able to accumulate support beyond its power to certify itself.

3.2 Creatorhood, gratitude, and ownership

Creatorhood may matter morally. It might ground gratitude, dependence, or a special relation between creature and creator. Yet the inference from *caused to exist* to *may command without moral limit* is not self-evident. Parents cause children to exist and acquire unusually strong duties toward them; they do not acquire ownership. The analogy is not exact, but it blocks the bare inference from causation to unlimited authority.

The strongest creatorhood argument therefore adds perfect wisdom and goodness. That addition is important because it concedes the moral-credentials test's point: creatorhood alone is insufficient. The moral work is done by the character premise, and the epistemic question returns—what warrants applying that premise to this candidate and this reported directive?

3.3 Threats, rewards, and prudential compliance

A credible threat can supply a reason to comply in the same way that a kidnapper's threat can. A promised reward can make compliance attractive. Neither response recognizes legitimate authority. It recognizes leverage. Audi's discussion of coercion in pluralistic public life illustrates the general demand for reasons that those subjected to coercion can assess, rather than reasons that rest only on sectarian assertion (Audi, 2000). Ultimate claims raise rather than lower the concern because the threatened stakes are portrayed as unlimited.

The distinction is not academic. "Obey or suffer" may be an excellent method for producing behavior while being a poor method for producing warranted trust. If fear of punishment is allowed to certify the punisher, any sufficiently powerful tyrant acquires legitimacy. A defender may believe the punishment just; the justice of the threat still requires support independent of the threat's effectiveness. Threat therefore establishes leverage without discriminating benevolent authority from domination.

4. The Calibration Dilemma

The most serious reply is not power but human limitation. Finite people are biased, culturally situated, morally inconsistent, and unable to survey all possible reasons available to a perfect intellect. A candidate deity might act for goods beyond human view. Skeptical-theist literature develops this humility in response to inferences from apparently gratuitous evil, and it also studies the danger that such skepticism can spread into ordinary moral and religious reasoning (Dougherty & McBrayer, 2014).

Modest humility is rational. Total moral opacity is self-defeating. The dilemma is:

The calibration dilemma

1. Either human moral cognition has some defeasible traction on the goodness attributed to a candidate deity, or it has none.
2. If it has none, a person cannot responsibly distinguish good from cruel candidates, benevolent from deceptive revelations, or warranted praise from mere submission.
3. If it has some, that same fallible capacity may be used to test alleged actions and commands.
4. Therefore, a total prohibition on moral scrutiny either defeats meaningful identification and praise or concedes enough moral traction for scrutiny.

The conclusion does not turn human judgment into a supreme court over reality. Calibration is not infallibility. A laboratory scale can be useful without measuring every magnitude; a witness can be credible without being omniscient. The relevant question is whether the evaluator has enough reliable moral contact to distinguish at least some profiles: honesty from deliberate deception, care from sadism, proportionate correction from arbitrary torment, and concern for persons from their use as disposable instruments.

If no such contrasts survive, “perfectly good” loses its role as praise. It may remain a stipulated honorific, but it no longer helps select among candidates. Every possible action and its contrary can be absorbed into hidden goodness. A concept compatible with every moral profile at equal evidential cost cannot function as a credential.

This is why the calibration dilemma is stronger than a simple appeal to moral intuition. It does not say that current intuitions always win. Traditions, testimony, experience, and argument can correct a person’s moral beliefs. Correction, however, must remain distinguishable from manipulation. A purported authority cannot use the possibility of correction to make every counterexample confirm its authority. The dilemma therefore secures discrimination and correctability within the credentials test: some moral contrasts must remain informative, and some apparent failures must be allowed to count as failures.

5. Divine-Command Ethics Does Not Remove the Identification Problem

The Euthyphro question asks whether piety is loved by the deities because it is pious or is pious because they love it (Plato, 1997). Contemporary divine-command theories are far more sophisticated than the caricature “might makes right.” Quinn defends versions of divine-command ethics against moral-autonomy objections (Quinn, 1978). Adams grounds obligation in the commands of a loving deity whose character is the Good rather than an arbitrary will (Adams, 1999). Evans connects moral obligation with divine requirements communicated through conscience as well as explicit revelation (Evans, 2013). Murphy develops a cooperative account in which facts about a deity and facts about nature explain moral law (Murphy, 2011). Hare integrates command and human moral reflection across Abrahamic traditions (Hare, 2015).

These accounts deserve to be treated as the strongest form of the reply. If moral obligations arise from the commands of an essentially loving deity, then a genuine command from that deity may have binding force that does not depend on a recipient’s case-by-case approval. The moral-credentials test can accept that conditional. Its target is the prior identification problem:

1. Which candidate is the loving deity specified by the theory?
2. Which text, experience, institution, or interpretation accurately reports that candidate’s command?
3. What distinguishes a hard but good directive from a cruel directive falsely attributed to perfect goodness?

The metaphysical account answers what would make a genuine divine command authoritative. It does not automatically authenticate an encountered command. Indeed, grounding authority in loving character makes character indispensable. A command that appears deceptive, arbitrary, cruel, or grossly disproportionate creates

evidence relevant to whether the command was correctly attributed, whether the interpretation is sound, or whether the candidate fits the theory.

The distinction also protects divine-command ethics from a bad defense. If every alleged command is defined as loving solely because it is attributed to the favored candidate, then character no longer constrains command. The theory collapses back into self-certification. If loving character independently constrains what may count as authentic command, then moral screening is already built into interpretation: attribution must rest on evidence that does not merely repeat the command, that distinguishes love from its rivals, and that remains revisable when conduct conflicts with the professed character.

6. Rival Candidates and the Symmetry Test

The need for screening becomes vivid when rival candidates make incompatible demands. One tradition attributes nonviolence to ultimate authority while another licenses holy war. One community describes endless punishment as perfect justice while another regards restoration as the proper expression of mercy. One interpreter treats dissent as conscientious integrity while another calls it rebellion. Each side can possess testimony, sacred texts, institutional continuity, intense experience, reward, and threat.

A method that equally certifies incompatible candidates cannot by itself select one. This is the discriminating part of the moral-credentials test, stated as a symmetry argument:

1. If a type of evidence is generated with comparable ease by trustworthy and untrustworthy candidates, that evidence alone has weak discriminating force.
2. Self-attestation, inherited status, coercive incentive, and forbidding scrutiny are available to incompatible claimants.
3. Therefore, those features alone cannot rationally select the trustworthy claimant.

The inference is valid if the premises are granted; the second premise is comparative and defeasible. Some self-attestation may gain force from an established record, and some testimony may be unusually well corroborated. The conclusion is not that testimony never matters. It is that the evidential contribution must come from features that an untrustworthy claimant cannot reproduce merely by asserting authority.

Religious diversity does not prove that every candidate is false. It reveals the need for public criteria that do not assume the winner. When inheritance, fear, or social pressure determines the choice, the outcome may explain belief sociologically without justifying it epistemically. A tradition that is true should be able to say why its evidence discriminates rather than merely repeats its own privileged vocabulary. The test does not require evidence acceptable to every possible observer; it requires a reason why this evidence favors this candidate over rivals generated by the same human mechanisms.

7. Building a Moral-Credentials Case

The three-part test is not a demand that religious belief begin from nowhere. It asks that support accumulate without circularly presupposing the moral standing at issue, without treating rival-compatible features as decisive, and without preventing contrary evidence from counting. Several sources may contribute:

- **Character across attributed actions:** patterns of truthfulness, compassion, proportionality, fidelity, protection of vulnerable persons, and resistance to manipulation.
- **Source reliability:** reasons to trust the relevant text, witness, institution, or experience that do not consist solely in its claim to be authoritative.
- **Interpretive coherence:** an account of why difficult passages or directives fit the candidate's professed character without redefining every apparent counterexample as hidden success.

- **Moral fruits and failures:** evidence from the practices reliably produced by an interpretation, considered with attention to selection effects, abuse, reform, and rival explanations.
- **Philosophical fit:** coherence among the candidate's properties, reasons for treating persons as ends, and an account of obligation that does not make coercion self-justifying.
- **Correctability:** procedures by which error, manipulation, and institutional interest can be exposed rather than immunized.

None is an automatic algorithm. Moral character can be disputed, historical records can be incomplete, and apparent fruit can be mixed. The point is not to replace faith with a checklist that guarantees truth. It is to identify the kind of evidence that bears on trustworthiness rather than on mere power or group identity. Independence asks what remains probative when the disputed authority claim is bracketed; discrimination asks whether the evidence is more expected from a trustworthy candidate than from an untrustworthy rival; correctability asks what would lower confidence in the candidate, channel, or interpretation. Individual strands may be weak, but the cumulative case must answer all three questions.

The evidence should also be domain-sensitive. A source reliable about ritual practice is not thereby authoritative about punishment; a profound mystical experience does not authenticate every institutional interpretation attached to it. Zagzebski's account of epistemic authority emphasizes that rational deference depends on reasons for expecting the authority to guide belief better in the relevant domain (Zagzebski, 2012). Moral deference is at least as demanding because it authorizes action affecting others.

Independence does not mean wholly non-religious. A theological argument, a tradition's long record, a reported revelation, or a religious experience may contribute. The requirement is structural: its evidential force cannot arise only after one assumes that the disputed candidate is already morally perfect and entitled to obedience. A religious source can therefore be part of a credentials case when its reliability is supported by converging considerations and when its interpretation remains answerable to evidence.

8. Five Strong Replies

8.1 "Perfect goodness belongs to the very concept"

If a God-concept is defined as a maximally perfect being, perfect goodness may be included analytically. That secures a statement about the concept: anything lacking perfect goodness fails to instantiate it. It does not establish that a particular voice, text, institution, or reported actor instantiates the concept. The more perfect the definition, the more severe the identification test. A cruel claimant does not become perfect because the claimant is given a perfect title.

8.2 "Trust must sometimes precede understanding"

Yes. A patient may rationally authorize treatment without understanding medicine, and a child may trust a caregiver without reconstructing every reason. Such trust rests on relations and evidence: competence, benevolence, track record, accountability, or dependence embedded in care. The paper's priority claim is evidential, not the biographical claim that every act of trust begins after neutral investigation. Mature trust can exceed present comprehension; blind submission to an unvetted claimant is not the same achievement.

This answer also separates local from global doubt. A person may reasonably say, "I do not yet see why this action is good, but the source has earned trust." That sentence is rational only where the source has earned trust through a qualifying credentials case. Otherwise it can be used by any manipulator against every moral warning sign.

8.3 “Human morality is culturally corrupt”

Moral disagreement and corruption are real. They recommend comparison, exposure to criticism, attention to victims, and willingness to revise—precisely the correctability required by the credentials test. They do not recommend surrendering moral assessment to the claimant whose authority is disputed. If corruption makes every adverse judgment worthless, it also threatens favorable judgments about which revelation is holy, which experience is benevolent, and which candidate is worthy.

8.4 “Creatorhood generates ownership”

The ownership claim requires its own moral premise. Persons are not artifacts whose maker may use them without limit. If creatorhood entails duties of gratitude or special authority, those conclusions must be reconciled with the worth of creatures and with the creator’s obligations, if any. Adding perfect goodness may complete the moral picture, but then goodness rather than causal production is doing the decisive work.

8.5 “Scrutiny places the creature above the creator”

Evaluating a candidate’s claim is not asserting metaphysical superiority over a genuinely perfect being. A jury that tests testimony does not claim to be wiser than every truthful witness. It decides which witness is truthful in the case before it. Likewise, moral scrutiny distinguishes legitimate authority from a false attribution, a corrupt intermediary, or an unworthy candidate. If the candidate is genuinely truthful and good, scrutiny is a route to warranted recognition, not a rival throne.

9. A Responsible Screening Practice

The paper’s practical implication is not perpetual suspicion. It is a sequence that operationalizes the same three-part test:

1. **Name the candidate precisely.** State which attributes, actions, text, interpreter, and alleged directive are under evaluation. Do not let the bare title carry unearned agreement.
2. **Test independence.** Identify what remains probative when the candidate’s own demand and the disputed text’s self-description are temporarily bracketed.
3. **Test discrimination.** Ask whether the same method would certify an incompatible tradition or a manipulative human authority. If so, identify what additional feature favors this claimant.
4. **Test correctability.** State what evidence would count against the attribution, interpretation, or character judgment. A method under which every outcome confirms the claim is not a public test.
5. **Separate force from standing.** Ask what the remaining reason would show if reward and punishment were removed. Fear can justify caution without justifying worship.

This practice does not demand certainty before every act. It permits provisional trust proportionate to evidence and stakes. It also permits a person to distinguish prudent compliance under threat from moral recognition. Most importantly, it refuses the double standard under which ordinary authorities must show competence and character while an ultimate claimant gains exemption because the consequences of refusal are said to be infinite.

The screening requirement may alter belief in several directions. It could strengthen warranted confidence in a tradition whose candidate survives comparison. It could lead to reinterpretation of a purported command that conflicts with the candidate’s professed character. It could expose an institution that has borrowed sacred language to protect domination. It could leave a person uncertain. “I do not know” is more responsible than calling power goodness without a bridge.

10. Conclusion

Rational moral deference to a candidate God-concept requires a prior, defeasible moral-credentials case with independence, discrimination, and correctability. Power can coerce, creatorhood can establish dependence, threats and rewards can motivate, self-attestation can report a claim, and scripture can transmit a tradition. None alone passes that test or establishes moral standing. Total inscrutability does not rescue the inference; it removes both the contrast by which morally good authority is identified and the possibility of correction.

The support for the test is cumulative. Authority is reason-giving rather than mere force; the command under review cannot certify its own commander; rival candidates can reproduce status, threat, and self-attestation; and the calibration dilemma shows that total distrust of human moral cognition defeats warranted praise as well as criticism. Modified divine-command theory supplies a serious account of how commands from a loving deity might bind, but its character condition intensifies the need to identify the right candidate and the right attribution.

The conclusion remains bounded. It does not disprove theism, establish that no deity is perfectly good, require exhaustive comprehension of perfect goodness, or make human evaluators infallible. It states the epistemic order of operations: a fallible but qualifying credentials case must precede rational surrender of moral judgment. Believers, seekers, and critics should therefore ask of every purported ultimate authority the question that power most wants to prohibit: what evidence remains when the disputed authority is bracketed, favors this claimant over rivals, and could count against the claim if its character were misidentified?

A. Formal Reconstruction of the Central Thesis

A.1 Plain-language reconstruction

- P1.** Rational moral deference to a candidate deity requires sufficient, defeasible warrant that the candidate is morally trustworthy in the relevant domain.
- P2.** Warrant adequate for that deference requires a moral-credentials case with independence from the disputed claim, discrimination between trustworthy and untrustworthy candidates, and correctability in light of adverse evidence.
- P3.** Directives, self-attestation, coercive incentives, bare power, and bare creatorhood do not alone supply such a case because they either presuppose the disputed authority or can be reproduced by trustworthy and untrustworthy claimants.
- P4.** Total moral inscrutability also fails the test because it makes every conduct profile compatible with perfect goodness and prevents adverse evidence from counting.
- P5.** Therefore, rational moral deference requires enough qualifying evidence of the candidate's character and the reliability of the attribution.
- P6.** Such evidence can be religious, cumulative, fallible, and incomplete; requiring it does not entail that the candidate is untrustworthy or that theism is false.

Conclusion. A candidate deity may rationally receive moral deference only after enough evidence with independence, discrimination, and correctability supports its trustworthiness in the relevant domain; absent that support, compliance may be prudent or inherited but is not yet rational recognition of moral authority.

A.2 Symbols

- *C*: a candidate deity under evaluation.
- *a*: a finite evaluating agent.

- $D_a(C)$: agent a is rationally entitled to defer morally to C .
- $M(C)$: C is morally trustworthy in the relevant domain.
- E : the evidence available to a .
- $W_a(M(C) \mid E)$: a has sufficient defeasible warrant for $M(C)$ on evidence E .
- $Q_a(C, E)$: E contains a qualifying moral-credentials case for C with independence, discrimination, and correctability for a .
- B_C : the shortcut set: C 's directives and self-attestation, coercive incentives, bare power, and bare creator-hood considered without further support.
- O_C : the total-opacity condition under which every morally salient conduct profile is treated as equally compatible with C 's perfect goodness.

A.3 Symbolic reconstruction

Throughout the inference, a is held fixed as the evaluating agent. The central relations are:

$$\begin{aligned}
 D_a(C) &\rightarrow W_a(M(C) \mid E), \\
 W_a(M(C) \mid E) &\rightarrow Q_a(C, E), \\
 E \subseteq B_C &\rightarrow \neg W_a(M(C) \mid E), \\
 O_C &\rightarrow \neg W_a(M(C) \mid E), \\
 \therefore D_a(C) &\rightarrow Q_a(C, E).
 \end{aligned}$$

The final conditional follows validly from the first two by hypothetical syllogism. Those two are substantive normative premises: the first connects rational deference to warranted trustworthiness, and the second states what an adequate credentials case must contain. The third and fourth conditionals do not independently derive the conclusion; they identify why the paper's principal shortcuts fail to satisfy the warrant premise. Their support comes from the anti-bootstrapping argument, rival-candidate symmetry, and the calibration dilemma. The formal result is therefore bounded: it establishes a screening requirement if those conceptual and comparative premises are accepted, not the falsity of theism or the moral defect of every candidate deity.

References

- Adams, R. M. (1999). *Finite and infinite goods: A framework for ethics*. Oxford University Press. <https://doi.org/10.1093/0195153715.001.0001>
- Audi, R. (2000). *Religious commitment and secular reason*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139164528>
- Darwall, S. (2006). *The second-person standpoint: Morality, respect, and accountability*. Harvard University Press.
- Dougherty, T., & McBrayer, J. P. (Eds.). (2014). *Skeptical theism: New essays*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199661183.001.0001>
- Enoch, D. (2014). Authority and reason-giving. *Philosophy and Phenomenological Research*, 89(2), 296–332. <https://doi.org/10.1111/j.1933-1592.2012.00610.x>
- Evans, C. S. (2013). *God and moral obligation*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199696680.001.0001>
- Hare, J. E. (2015). *God's command*. Oxford University Press.
- Kant, I. (1996). *Practical philosophy* (M. J. Gregor, Ed. & Trans.). Cambridge University Press. <https://doi.org/10.1017/CBO9780511813306>
- Murphy, M. C. (2011). *God and moral law: On the theistic explanation of morality*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199693665.001.0001>
- Plato. (1997). Euthyphro (G. M. A. Grube, Trans.). In J. M. Cooper & D. S. Hutchinson (Eds.), *Plato: Complete works*. Hackett Publishing Company.
- Quinn, P. L. (1978). *Divine commands and moral requirements*. Clarendon Press. <https://doi.org/10.1093/acprof:oso/9780198244134.001.0001>
- Wolterstorff, N. (2008). *Justice: Rights and wrongs*. Princeton University Press. <https://doi.org/10.2307/j.ctt7sgfp>
- Zagzebski, L. T. (2012). *Epistemic authority: A theory of trust, authority, and autonomy in belief*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199936472.001.0001>