

The Architecture of Moral Inconsistency A Philosophical Investigation of Contextual Moral Valuation

Momen Ghazouani *Editor-in-Chief, The Atlantic Journal*

April 4, 2026

Abstract

Human moral judgment exhibits a puzzling feature: we routinely condemn harmful actions in principle while simultaneously expressing sympathy toward those who commit such actions under specific circumstances. This paper investigates whether this apparent inconsistency reveals a deep structure in moral cognition or constitutes genuine moral failure. I develop a philosophical framework *Contextual Moral Valuation* (CMV) that situates this phenomenon at the intersection of normative ethics, moral psychology, and metaethics. The framework posits that moral evaluation emerges from the interaction of reward valuation, threat assessment, cognitive bias, and psychological distance. I argue that this structure generates what I call the *Contextual Moral Alignment Paradox* (CMAP) : the simultaneous condemnation and endorsement of the same agent under different contextual parameters. This paper makes three central contributions. First, I demonstrate that CMAP cannot be dismissed as simple hypocrisy or cognitive error it reflects fundamental features of practical reason operating under conditions of value pluralism and uncertainty. Second, I explore the normative implications: whether context-sensitive moral judgment can be rationally defensible or whether it always constitutes moral failure. Third, I examine how CMV challenges traditional assumptions in moral philosophy about consistency, impartiality, and the relationship between moral principles and moral judgment. The analysis reveals that what appears as moral inconsistency may instead represent a rational response to genuine moral complexity, though one that remains vulnerable to systematic distortion and motivated reasoning.

Keywords: moral judgment, moral consistency, psychological distance, moral contextualism, practical reason, value pluralism, moral psychology, normative ethics

1. Introduction: The Problem of Selective Moral Alignment

1.1 The Puzzle

Consider the following cases :

Case 1 (The Benefactor): Citizens of Country A condemn human rights violations in principle. Yet when their nation’s leader implements policies that violate rights abroad but generate substantial domestic economic benefits, many express continued support for this leader, viewing them as “doing what’s necessary” or “putting our interests first.”

Case 2 (The Corporation): An employee morally condemns child labor practices. Yet when their own company is discovered using suppliers who employ children in distant countries a practice that helps maintain the employee’s salary and benefits they express understanding of “market pressures” and “competitive realities” rather than moral outrage.

Case 3 (The Historical Figure): A scholar vigorously opposes slavery and considers it an absolute moral wrong. Yet this same scholar can discuss historical slave-owning figures with admiration for their other contributions, can teach their works without moral condemnation, and can even express a form of selective sympathy when considering the “constraints of their time.”

These cases share a common structure: an agent simultaneously holds that certain actions are morally wrong *in principle* while expressing positive evaluation, sympathy, or support toward *particular agents* who commit such actions under specific conditions. This pattern appears to violate a basic requirement of moral rationality consistency yet it is ubiquitous in human moral judgment.

1.2 Why This Matters Philosophically

The phenomenon of selective moral alignment poses several deep philosophical problems :

The Consistency Problem: If moral principles are genuinely universal and action-guiding, how can rational agents endorse both the principle “X is wrong” and the particular judgment “Agent A’s doing X is acceptable (or even admirable)”? This appears to violate the universalizability constraint that many philosophers consider essential to moral reasoning.

The Rationality Problem: Is selective moral alignment a form of practical irrationality a failure to properly apply one’s own moral commitments? Or might it reflect a sophisticated response to genuine moral complexity that our theoretical frameworks inadequately capture?

The Normative Problem: Even if we can explain *why* humans exhibit selective moral alignment (through psychology, neuroscience, or evolutionary history), does this explanation have normative force? Should we revise our moral

theories to accommodate this phenomenon, or should we regard it as a persistent form of moral failure to be corrected?

The Metaethical Problem: What does the ubiquity of context-dependent moral judgment tell us about the nature of moral truth, moral knowledge, and moral motivation? Does it support moral relativism, or can it be reconciled with moral objectivism?

1.3 Existing Approaches and Their Limitations

Contemporary moral philosophy offers several frameworks that address aspects of moral context-sensitivity, but none fully captures the phenomenon I am investigating.

Particularism (Dancy 2004) holds that moral principles are at best summaries of past cases rather than universal laws. While particularism accommodates context-sensitivity, it does not explain why context affects judgment in *systematic and predictable ways* related to self-interest and psychological distance rather than in ways purely responsive to morally relevant features.

Virtue Ethics (Aristotle; MacIntyre 1981) emphasizes character and practical wisdom over universal principles. Yet virtue ethics typically treats apparent inconsistency as vice (e.g., self-deception, weakness of will) rather than as potentially revelatory of moral structure.

Moral Psychology (Greene 2013; Haidt 2012) has documented extensive context-sensitivity in moral judgment but largely focuses on descriptive explanation rather than normative evaluation. These approaches leave open the question of whether context-sensitive judgment constitutes good or bad moral reasoning.

Consequentialism and **Deontology**, in their standard forms, provide little room for the kind of systematic context-sensitivity we observe. They typically treat such patterns as reflecting either (1) implicit consequentialist calculation that the agent fails to recognize, or (2) moral failure due to self-interest corruption.

What is needed is a framework that takes seriously both the *systematic nature* of context-dependent moral judgment and the *normative question* of whether such judgment can be rationally defensible.

1.4 The Contextual Moral Valuation Framework

I propose the Contextual Moral Valuation (CMV) framework as a philosophical model that integrates empirical observations about moral judgment with normative analysis. The framework rests on four key claims:

1. **Structural Claim:** Moral judgment emerges from the integration of multiple evaluative dimensions reward valuation, threat assessment, bias mod-

ulation, and psychological distance rather than from the straightforward application of universal principles.

2. **Psychological Claim:** These dimensions interact in systematic ways that produce the Contextual Moral Alignment Paradox (CMAP) simultaneous condemnation and endorsement of the same agent.
3. **Normative Claim:** CMAP is not always irrational or morally indefensible, but neither is it always rationally justified. Whether particular instances constitute good or bad moral reasoning depends on complex factors that traditional moral theories inadequately address.
4. **Metaethical Claim:** The structure of CMV suggests that moral judgment operates as a form of *situated practical reason* that is sensitive to the agent's epistemic, motivational, and relational position rather than as the detached application of universal principles.

This paper develops these claims systematically, examines their philosophical implications, and considers objections. My goal is not merely to describe how humans make moral judgments but to investigate what this structure reveals about the nature of moral thought and whether it can be normatively defended.

2. The Architecture of Moral Evaluation: Components and Integration

To understand how selective moral alignment arises, we must first examine the structure of moral evaluation itself. I propose that moral judgment integrates four distinct evaluative dimensions.

2.1 Reward Valuation (R) : The Assessment of Benefit

The first component is the evaluation of potential benefits associated with the moral agent or their actions. This includes:

- **Direct benefits:** Material goods, resources, opportunities that the agent provides to the evaluator
- **Indirect benefits:** Social goods, collective benefits to the evaluator's group or community
- **Symbolic benefits:** Psychological goods such as identity affirmation, meaning, or status

Philosophical Significance: The inclusion of reward valuation in moral judgment might seem to threaten the very idea of morality as impartial or principled. Yet three considerations suggest that reward sensitivity is not inherently incompatible with genuine moral judgment:

First, even impartial moral theories recognize that benefits and harms matter morally. Consequentialism makes this explicit; even deontology incorporates

beneficence as a duty. The question is not whether benefits are morally relevant but how they should be weighted relative to other considerations.

Second, epistemic humility suggests that agents who benefit from another's actions may have access to morally relevant information (about intentions, consequences, or contextual constraints) that distant observers lack. Benefit is not merely a bias but potentially an informational signal.

Third, from a virtue ethics perspective, gratitude and loyalty both of which involve positive evaluation of benefactors are often considered virtues rather than vices. The complete suppression of positive feeling toward those who benefit us might reflect moral deficiency rather than moral purity.

The challenge, then, is not to eliminate reward valuation from moral judgment but to understand when and how it appropriately influences evaluation.

2.2 Threat Detection (T): The Assessment of Harm

The second component is the evaluation of potential harms, dangers, or moral transgressions associated with the agent. This includes:

- **Direct harm:** Physical, psychological, or material damage caused by the agent
- **Moral transgression:** Violations of rights, duties, or moral principles
- **Systemic threat:** Dangers to social institutions, norms, or collective goods

Philosophical Significance: Harm and wrongdoing are central to nearly all moral theories. The puzzle is not why threat matters but why threat sensitivity varies systematically across contexts in ways that seem morally arbitrary.

Classical moral philosophy typically assumes that moral wrongness should be evaluated independently of the evaluator's relationship to the agent or the harm. A murder is equally wrong whether committed by a stranger or a benefactor, whether the victim is near or far. Yet empirical evidence and phenomenological reflection suggest that moral response to wrongdoing is systematically modulated by contextual factors. This raises a deep question: Is stable threat response across all contexts a requirement of moral rationality? Or might variable threat response reflect appropriate sensitivity to epistemically and morally relevant contextual features?

2.3 Bias Modulation (B): Cognitive Reframing

The third component comprises various cognitive mechanisms that reframe, reinterpret, or reconstruct moral information. These include :

- **Motivated reasoning:** Selective attention to information that supports desired conclusions

- **Moral disengagement:** Cognitive restructuring that makes harmful actions seem acceptable (e.g., euphemistic labeling, diffusion of responsibility, dehumanization of victims)
- **System justification:** Tendency to rationalize existing social arrangements
- **Attribution asymmetries:** Interpreting the same behavior differently based on who performs it

Philosophical Significance: Bias modulation is perhaps the most philosophically troubling component. Unlike reward and threat, which involve substantive moral considerations (benefits and harms), bias modulation seems to involve purely epistemically and morally distorting processes.

Yet the philosophical status of bias modulation is more complex than it first appears. Consider three complications :

First, what we label “bias” may sometimes reflect legitimate moral considerations that dominant moral theories fail to recognize. For instance, the tendency to evaluate in-group members more favorably might reflect genuine moral obligations of special relationship and loyalty rather than mere bias.

Second, some forms of “bias” may be epistemically necessary given our limited cognitive capacities. We cannot evaluate all moral claims with equal scrutiny; we must use heuristics, and these inevitably produce systematic patterns that might be labeled “bias.”

Third, moral theories themselves may contain biases. The demand for perfect impartiality, for instance, might reflect a particular philosophical tradition rather than a universal requirement of rationality.

Nonetheless, many forms of bias modulation particularly motivated reasoning in service of self-interest clearly constitute epistemic and moral failures. The challenge is distinguishing legitimate contextual sensitivity from illegitimate bias.

2.4 Psychological Distance (D) : The Geometry of Moral Space

The fourth component is psychological distance the perceived separation between the evaluator and the consequences of the agent’s actions. This includes:

- **Spatial distance:** Geographic separation
- **Temporal distance:** Separation in time (past or future)
- **Social distance:** Degree of social connection or similarity
- **Hypothetical distance:** Whether consequences are certain or merely possible

Philosophical Significance: Psychological distance occupies a unique position in the CMV framework. Unlike the other components, distance is not itself a form of evaluation but rather a parameter that modulates how other evaluations function.

The philosophical puzzle is this: Should psychological distance affect moral judgment at all ?

The Impartialist Answer : Most modern moral philosophy, following Kant and Bentham, answers “No.” Moral principles should apply equally regardless of distance. A harm is equally wrong whether inflicted on someone near or far, now or later, similar or different.

The Particularist Answer: Some philosophers, particularly virtue ethicists and care ethicists, answer “yes.” Our moral obligations and appropriate responses differ based on relationships and proximity. We have special duties to those near to us.

The CMV Answer: The truth lies between these positions. Some distance effects reflect appropriate moral sensitivity (e.g., special obligations to those near), but many distance effects reflect epistemically and morally arbitrary features of human psychology that should be corrected.

Crucially, the CMV framework predicts *asymmetric* distance effects: psychological distance attenuates threat perception more than reward perception. This asymmetry has profound normative implications, as we shall see.

2.5 Integration : The Structure of Moral Judgment

How do these components combine to produce moral judgments? I propose that moral evaluation (M) emerges from a function integrating these dimensions :

$$M(\text{agent, context, subject}) = f[\mathbf{R}(\text{agent, context}), \mathbf{T}(\text{agent, context}), \mathbf{B}(\text{agent, context, subject}), \mathbf{D}(\text{agent, context})]$$

Where : - **f** is an integration function (potentially non-linear) - **R**, **T** are reward and threat valuations - **B** is bias modulation (positive values favor the agent, negative values disfavor) - **D** is psychological distance (with asymmetric effects on R and T)

For analytical purposes, consider a simplified form :

$$M = [\alpha \cdot R - \beta \cdot T] / D^{\gamma} + \delta \cdot B$$

Where α , β capture individual sensitivity to reward and threat, γ captures the strength of distance effects, and δ captures susceptibility to bias.

Philosophical Implications of the Integration Structure:

1. **Moral judgment is inherently multidimensional.** It cannot be reduced to a single principle (maximize utility, follow duty, express virtue) but reflects the integration of multiple evaluative considerations.
2. **Context affects not just *what* is evaluated but *how* evaluation occurs.** The same action by the same agent can receive different moral evaluations not merely because different facts are salient but because the evaluative architecture itself responds differently.

3. **Individual differences in moral judgment reflect not just different principles but different integration parameters.** Two people might agree on all morally relevant facts yet reach different conclusions because they weigh reward vs. threat differently or differ in susceptibility to bias.
4. **Moral judgment is path-dependent.** The order in which information is received, the framing of choices, and other contextual features can affect final evaluation even when all information is eventually available.

This integration structure generates the central phenomenon I aim to explain: the Contextual Moral Alignment Paradox.

3. The Contextual Moral Alignment Paradox (CMAP)

3.1 Defining the Paradox

The Contextual Moral Alignment Paradox occurs when an agent simultaneously:

1. **Endorses a moral principle** condemning certain actions (e.g., “torture is wrong,” “exploiting workers is immoral”)
2. **Expresses positive moral evaluation or sympathy** toward a particular agent who commits such actions under specific conditions

This is paradoxical because it appears to violate the universalizability of moral principles: if torture is wrong, it should be wrong regardless of who does it or what benefits it produces.

3.2 Conditions for CMAP

According to the CMV framework, CMAP emerges when :

$$R(\text{agent, context}) > [/] \cdot T(\text{agent, context}) - [D /] \cdot \cdot B(\text{agent, context, subject})$$

This inequality is most likely satisfied when: - **High Reward:** The agent provides substantial benefits to the evaluator - **High Distance:** The harms occur far away (spatially, temporally, or socially) - **Positive Bias:** Cognitive biases favor the agent (in-group membership, charismatic leadership, etc.) - **Individual Parameters:** The subject weights rewards more than threats (high / ratio)

3.3 Varieties of CMAP

CMAP manifests in several distinct forms:

Type 1: The Distant Harm Paradox

Condemn a harm in principle but tolerate agents who cause that harm to distant others while benefiting the evaluator.

Example: Condemning sweatshop labor while supporting companies that use it.

Type 2: The Historical Exception Paradox

Condemn a practice as morally wrong but express admiration for historical figures who engaged in it.

Example: Condemning slavery while celebrating slave-owning founding fathers.

Type 3: The Necessary Evil Paradox

Maintain that an action is morally wrong yet support agents who commit it for perceived greater goods.

Example: Believing torture is wrong but supporting leaders who authorize it for national security.

Type 4: The In-Group Exemption Paradox

Apply different standards to in-group vs. out-group members who commit the same actions.

Example: Harshly judging corruption by foreign leaders while excusing it in one's own politicians.

3.4 Is CMAP Genuinely Paradoxical?

One might object that CMAP is not a real paradox but merely reflects one of three possibilities:

Objection 1: Implicit Consequentialism

Perhaps the agent is simply an implicit consequentialist who recognizes that while an action type is generally wrong, particular tokens can be justified when consequences are sufficiently good. The apparent paradox dissolves because the agent's *real* principle is consequentialist, not categorical.

Reply: This objection fails for two reasons. First, empirical evidence suggests that people do not consistently apply consequentialist reasoning they show inconsistent trade-offs between harms and benefits depending on framing and context. Second, the pattern of CMAP does not track what consequentialist calculation would recommend (which should be invariant to psychological distance) but instead tracks psychologically arbitrary factors like spatial distance and in-group membership.

Objection 2: Purely Instrumental Evaluation

Perhaps positive evaluation of harmful-but-beneficial agents reflects purely pragmatic, non-moral judgment. The agent is not making a moral claim but merely expressing that "this person is useful to me."

Reply: While this explains some cases, phenomenological evidence suggests that CMAP often involves genuinely moral evaluation expressions of moral character ("they're a good leader"), moral justification ("they did what was necessary"), and moral sympathy ("I understand why they had to do it") rather than purely instrumental assessment.

Objection 3: Incomplete Information

Perhaps what appears as inconsistency simply reflects learning new information

that changes the moral evaluation.

Reply : While some cases involve information change, CMAP occurs even when all factual information is held constant and only contextual parameters (distance, benefit) vary. Moreover, CMAP predicts systematic patterns (distance effects, benefit effects) that are not explained by information quality.

I conclude that CMAP constitutes a genuine puzzle requiring explanation. The question is whether it reveals a defect in human moral reasoning or a feature that can be normatively defended.

4. The Rationality of Context-Dependent Moral Judgment

Having established that CMAP exists and is genuinely puzzling, I now turn to the normative question: Is context-dependent moral judgment rational? Can it be morally defensible?

4.1 The Case Against: CMAP as Moral Failure

The dominant view in moral philosophy would regard CMAP as clear evidence of moral failure. This view rests on several arguments:

Argument 1: Violation of Universalizability

Kant's universalizability principle holds that a maxim is morally acceptable only if it can be consistently willed as a universal law. CMAP violates this: one cannot consistently will both "torture is wrong" and "torture is acceptable when it benefits me or when victims are distant."

Argument 2: Arbitrary Discrimination

Moral evaluation should depend on morally relevant features (intentions, consequences, rights violations) not on morally arbitrary features (psychological distance, personal benefit). CMAP allows morally arbitrary factors to influence judgment.

Argument 3: Motivated Reasoning

Much of CMAP clearly results from self-serving bias and motivated reasoning epistemically deficient processes that we should correct rather than rationalize.

Argument 4: Moral Progress

Historical moral progress often involves expanding our moral circle and overcoming parochial attachments. CMAP represents precisely the kind of limited moral vision we should transcend.

These arguments are powerful. Yet I believe they do not settle the question.

4.2 The Case For: CMAP as Rational Response to Complexity

Several considerations suggest that at least some forms of context-dependent moral judgment might be rationally defensible:

Consideration 1: Epistemic Constraints and Rational Heuristics

Perfect moral rationality would require evaluating all moral claims with equal scrutiny, processing infinite information, and resisting all cognitive biases. But we are bounded agents with limited cognitive resources.

Psychological distance might serve as a *rational heuristic* for approximating morally relevant features: - **Epistemic access:** We typically have better information about near events than distant ones - **Causal influence:** We typically have more control over near outcomes than distant ones - **Relational obligations:** Many philosophers argue we have special duties to those near to us

If psychological distance correlates with these morally relevant features, distance-sensitive judgment might approximate good moral reasoning under resource constraints.

Counterpoint: Even if distance sometimes correlates with morally relevant features, the CMAP pattern shows *systematic distortion* distance effects that go beyond what epistemic access or special obligations would justify. The asymmetry (distance attenuates threats more than rewards) particularly suggests motivated reasoning rather than rational heuristics.

Consideration 2: Value Pluralism and Moral Complexity

Bernard Williams (1981) and other pluralists argue that moral life involves irreducibly plural values that sometimes conflict. When values conflict, trade-offs are necessary, and reasonable people might weigh values differently.

Perhaps CMAP reflects rational navigation of genuine value conflicts: - **Loyalty vs. Justice:** Special obligations to compatriots vs. universal duties to all humans - **Gratitude vs. Principle:** Acknowledging benefits received vs. condemning wrongs committed - **Practical Necessity vs. Moral Ideals:** What must be done given constraints vs. what would be ideal

On this view, the apparent inconsistency of CMAP reflects not irrationality but the absence of algorithmic solutions to genuine moral complexity.

Counterpoint: While value pluralism explains some context-sensitivity, it does not explain why context affects judgment in ways that systematically favor the self and the near. Value conflict might justify different weightings, but not weightings that consistently serve self-interest.

Consideration 3: The Agent-Relative Structure of Practical Reason

Some philosophers (Nagel 1986; Scheffler 1982) argue that practical reason is inherently agent-relative. We are not mere conduits for impersonal value but particular agents with particular projects, relationships, and perspectives. From this view, it is rationally appropriate for my evaluation of agents to be sensitive to how they relate to my projects and relationships. The fact that someone benefits me is not morally irrelevant but part of the morally significant web of relationships and reciprocities that constitute my practical life.

Counterpoint: Agent-relative permissions (I may give special weight to my projects) differ from agent-relative *moral evaluation*. I might be permitted to pursue my good even when it conflicts with others' goods, but this does not entail that I should *morally evaluate* wrongdoing differently based on whether it serves my interests.

4.3 A Nuanced Position: Distinguishing Defensible from Indefensible Context-Sensitivity

The arguments on both sides have force. I propose a more nuanced position :

Some context-sensitivity is rationally defensible: - Giving weight to special relationships and near obligations - Using psychological proximity as a rough proxy for epistemic access - Recognizing genuine value pluralism and trade-offs - Acknowledging agent-relative structure of practical reason

Much of CMAP, however, constitutes moral failure: - Systematic bias favoring personal benefit - Asymmetric distance effects (threats attenuated more than rewards) - Motivated reasoning that distorts rather than tracks moral reality - In-group favoritism that goes beyond special obligations

The Normative Task: The challenge is not to eliminate all context-sensitivity (which may be impossible and perhaps undesirable) but to distinguish legitimate from illegitimate forms and to develop strategies for mitigating distortions.

This requires : 1. **Reflective equilibrium:** Bringing our particular judgments into coherence with our considered principles 2. **Perspective-taking:** Actively imagining how situations appear from other viewpoints 3. **Bias awareness:** Recognizing our susceptibility to motivated reasoning 4. **Moral dialogue:** Engaging with others who occupy different positions in moral space

5. Metaethical Implications: What CMAP Reveals About Moral Truth

The phenomenon of CMAP and the structure of CMV have important implications for metaethics the philosophical investigation of the nature of moral truth, moral knowledge, and moral motivation.

5.1 The Challenge to Moral Objectivism

Robust moral objectivism holds that moral truths are mind-independent facts that rational agents should converge upon through proper reasoning. CMAP poses a challenge to this view:

The Convergence Problem: If moral truth is objective and accessible through reason, why do rational agents systematically diverge in their moral judgments based on factors (psychological distance, personal benefit) that seem morally irrelevant?

The Motivational Problem: Objectivism typically assumes that moral judgments have intrinsic motivational force (“moral internalism”). Yet CMAP shows that the same person can judge an action wrong yet feel motivated to support the agent who commits it suggesting that judgment and motivation come apart.

The Evolutionary Debunking Argument: If our moral judgment mechanisms evolved to serve fitness rather than track moral truth, and if CMAP reflects those evolved mechanisms, should we trust our moral judgments at all?

5.2 Does CMAP Support Relativism ?

One might conclude that CMAP supports moral relativism: moral truth varies with the evaluator’s position, relationships, and interests.

However, this conclusion is too quick. CMAP is compatible with several metaethical positions :

Sophisticated Objectivism: Moral truths are objective but complex. What appears as context-sensitivity might reflect : - Agent-relative moral truths (different obligations for differently situated agents) - Context-dependent applications of universal principles - Epistemically rational responses to uncertainty about objective values

Error Theory: All moral judgments are systematically false, and CMAP reveals the psychological mechanisms that generate these false beliefs. Moral discourse serves social coordination functions rather than representing moral reality.

Expressivism: Moral judgments express attitudes rather than beliefs about facts. CMAP shows how these attitudes are shaped by psychological factors, but this neither supports nor undermines the objectivity of values.

Constructivism: Moral truth is constructed through ideal deliberation or social agreement. CMAP reveals features of actual human moral psychology that ideal deliberation might correct or incorporate.

5.3 The Situated Rationality View

I propose a metaethical position that takes seriously both the reality of moral truth and the contextual structure of moral judgment:

Situated Rationality: Moral judgment operates as a form of *situated practical reason* reason that is necessarily exercised from a particular position in social and evaluative space, with particular epistemic and motivational constraints.

Key features : 1. **Positionality:** All moral judgment occurs from a position you cannot evaluate from nowhere 2. **Epistemic Boundedness:** We have better knowledge of near events; this legitimately affects confidence in moral judgment 3. **Motivational Structure:** Human moral motivation is structured by relationships and benefits; this is a feature of practical reason, not a bug 4.

Normative Aspiration: Despite situatedness, we can aspire toward more impartial judgment through reflective correction

On this view: - **Moral truths** are objective (or intersubjectively valid) - **Moral judgment** is necessarily situated and context-sensitive - **Moral progress** involves developing capacities to transcend situational biases while acknowledging that complete transcendence is neither possible nor necessarily desirable - **CMAP** reveals both legitimate features of situated judgment and systematic distortions requiring correction

5.4 Implications for Moral Knowledge

If moral judgment is structured as CMV suggests, what are the implications for moral knowledge?

Implication 1: Moral Knowledge is Perspectival but Not Merely Subjective

We can have genuine moral knowledge while acknowledging that such knowledge is necessarily acquired from a perspective. The evaluator's position provides epistemic access to some moral facts (about relationships, consequences for the near) while limiting access to others (about distant harms).

Implication 2: Self-Knowledge is Essential to Moral Knowledge

Understanding our own psychological tendencies (to favor benefactors, to discount distant harms) is crucial for correcting distortions. Moral knowledge requires not just knowledge of moral principles but knowledge of how our judgment mechanisms operate.

Implication 3: Moral Testimony and Multiple Perspectives are Essential

If individual judgment is systematically biased by position, moral knowledge requires integrating multiple perspectives. Those who benefit from an agent will systematically see them differently than those who are harmed neither perspective alone provides complete moral truth.

Implication 4: Moral Uncertainty is Pervasive

The CMV framework suggests that moral certainty is often epistemically inappropriate. Given our susceptibility to bias and our limited perspectives, we should often hold moral judgments with uncertainty and be open to revision.

6. Philosophical Objections and Replies

6.1 The Kantian Objection: Violation of the Categorical Imperative

Objection: CMV and CMAP directly contradict Kant's moral philosophy. The Categorical Imperative requires that we act only on maxims we could will as universal law and that we treat humanity always as an end, never merely as

a means. CMAP violates both formulations by allowing personal benefit and psychological distance to determine moral evaluation.

Reply : This objection is powerful but admits several responses:

First, the CMV framework is descriptive and diagnostic, not prescriptive. It describes how humans actually make moral judgments, not how they should. One can accept CMV as accurate description while adopting Kantian norms as prescriptive ideals.

Second, even Kant recognized imperfect duties that allow latitude in application and acknowledged that we have special obligations based on relationships. A Kantian can incorporate some context-sensitivity without abandoning the core Kantian project.

Third, the situated rationality view suggests that perfect impartiality may be psychologically impossible for embodied, relational agents like humans. The question then becomes: given our cognitive architecture, what is the best we can do? This is compatible with endorsing Kantian principles as regulative ideals even if perfect compliance is unattainable.

6.2 The Consequentialist Objection: Arbitrary Weighing of Outcomes

Objection : From a consequentialist perspective, CMAP looks like irrational inconsistency. All that matters morally is the total consequences for well-being. Psychological distance is morally irrelevant a harm to someone far away is exactly as bad as the same harm to someone near. Personal benefit is relevant only insofar as it contributes to overall utility. CMAP involves giving arbitrary weight to morally irrelevant factors.

Reply:

First, consequentialists face their own challenge from context-sensitivity. Empirical evidence suggests that even committed consequentialists exhibit distance effects and benefit biases in their actual judgments. Either we dismiss all such judgments as irrational (implausibly) or we acknowledge that bounded rationality requires heuristics that produce systematic deviations from ideal consequentialism.

Second, some consequentialists (satisficing consequentialism, scalar consequentialism) allow that agents need not always maximize impartially. Perhaps context-sensitive heuristics are acceptable if they produce good-enough outcomes.

Third, the situated rationality view can incorporate consequentialist concerns: we should work to reduce arbitrary distance effects precisely because they lead to worse outcomes by causing us to ignore distant suffering.

6.3 The Virtue Ethics Objection: Character Over Context

Objection: Virtue ethics emphasizes stable character traits rather than context-sensitive judgments. A virtuous person would not exhibit CMAP because virtue involves stable dispositions to respond appropriately across contexts. CMAP reveals vice (e.g., self-deception, cowardice, disloyalty to moral truth) rather than anything rationally defensible.

Reply :

First, even Aristotle recognized that virtue requires practical wisdom (*phronesis*) the ability to judge what is appropriate in particular circumstances. This involves context-sensitivity, not blind application of rules.

Second, virtue ethics can acknowledge tragic situations where virtues conflict (loyalty vs. justice, gratitude vs. truth-telling). CMAP might sometimes reflect such tragic conflicts rather than simple vice.

Third, if CMAP is a deep feature of human moral psychology, then cultivating virtue may require developing metacognitive capacities to recognize and correct for these tendencies rather than pretending they don't exist.

6.4 The Feminist Ethics Objection : Ignoring Relational Context

Objection: Feminist care ethics and relational ethics emphasize that moral judgment should be sensitive to relationships, context, and particular needs. From this perspective, traditional moral philosophy's demand for impartiality is itself problematic. CMV might vindicate context-sensitivity as appropriate moral response.

Reply :

I am sympathetic to this objection. Care ethics provides important correctives to impartialist ethics. However, three points are essential:

First, care ethics emphasizes responsiveness to actual needs and relationships, not to personal benefit. Much of CMAP reflects self-interest rather than caring attention to others.

Second, care ethics typically emphasizes responding to vulnerability and need, which should make distant victims more rather than less morally salient. The asymmetric distance effects in CMAP go against care ethics commitments.

Third, while care ethics values relationships, it does not endorse unlimited partiality. Even care ethicists recognize the moral pull of impartial concerns and the need to prevent caring relationships from becoming vehicles for domination or injustice.

6.5 The Evolutionary Debunking Objection

Objection: If CMAP reflects evolved psychological mechanisms designed to maximize fitness rather than track moral truth, we have reason to distrust it. This is a version of the evolutionary debunking argument: our moral judgments are unreliable because they were selected for reproductive success, not moral accuracy.

Reply:

First, the debunking argument proves too much. If we distrust all evolved cognitive mechanisms, we must distrust logic, perception, and rational inference as well all evolved capacities.

Second, evolution can produce reliable tracking mechanisms when tracking truth had fitness benefits. Our perceptual systems are reliable precisely because seeing accurately was fitness-enhancing. Perhaps the same applies to some aspects of moral judgment.

Third, even if evolution explains the origin of our moral capacities, it doesn't follow that these capacities only track fitness. Evolution gave us capacities for mathematics and philosophy that we now use for purposes unrelated to fitness.

Fourth, the CMV framework itself provides resources for correction. By understanding the systematic biases our evolved psychology produces, we can develop strategies to mitigate them much as we use telescopes to overcome limitations of evolved vision.

7. Practical Implications: Moral Reform and Moral Progress

Understanding CMAP is not merely of theoretical interest. It has important practical implications for how we should think about moral education, moral reform, and moral progress.

7.1 Implications for Moral Education

Implication 1: Teach Metacognitive Awareness

Moral education should include explicit instruction about psychological biases, motivated reasoning, and distance effects. Students should learn to recognize when their moral judgments might be distorted by self-interest or psychological distance.

Implication 2: Practice Perspective-Taking

Since our natural tendency is to view situations from our own position, moral education should systematically practice adopting other perspectives especially the perspectives of those who are harmed, distant, or different.

Implication 3: Develop Reflective Equilibrium Skills

Students should learn to bring particular judgments into coherence with general

principles through reflective iteration, recognizing inconsistencies and working to resolve them.

Implication 4: Acknowledge Moral Difficulty

Rather than presenting morality as simple rule-following, education should acknowledge the genuine difficulty of moral judgment under conditions of uncertainty, value pluralism, and cognitive limitation.

7.2 Implications for Political Discourse

Implication 1: Recognize Structural Bias in Evaluation

Political discourse should explicitly acknowledge that citizens will tend to evaluate leaders more favorably when those leaders provide benefits, even if they commit wrongs. This is not pure hypocrisy but a systematic feature of moral psychology requiring institutional correction.

Implication 2: Reduce Psychological Distance

Media, education, and political communication should work to reduce psychological distance between citizens and distant victims of policy. This includes vivid narratives, personal testimonies, and direct encounters.

Implication 3: Institutional Design

Since individual moral judgment is systematically biased, institutional design should incorporate checks and balances that don't rely on impartial individual judgment. This includes independent oversight, transparent decision-making, and accountability mechanisms.

7.3 Implications for Moral Progress

The Moral Arc Question: Does moral history show progress? Philosophers disagree, but if we accept that there has been moral progress (expansion of rights, reduction of violence, extension of moral concern), understanding CMAP helps explain both progress and resistance to it.

How CMAP Explains Moral Progress: - Reduction in psychological distance through technology, travel, and communication makes distant others more morally salient - Institutional reforms (human rights law, international criminal courts) create accountability mechanisms that overcome individual bias - Moral movements consciously work to reduce distance through narrative and testimony

How CMAP Explains Resistance to Progress: - Each expansion of moral concern threatens existing benefits to privileged groups - Psychological distance allows continued exploitation of distant others - Motivated reasoning generates rationalizations for resisting moral expansion

Strategic Insight: Effective moral reform requires addressing not just principles but the psychological mechanisms that generate resistance. This means: - Making distant harms more psychologically proximate - Disrupting

motivated reasoning through evidence and accountability - Creating conditions where moral principle and self-interest align rather than conflict

7.4 Implications for Individual Moral Life

Implication 1: Practice Moral Humility

Recognizing our susceptibility to CMAP should induce humility about our moral judgments, especially when we benefit from an agent's actions or when harms are distant.

Implication 2: Develop Corrective Practices

Individuals can develop practices to counter CMAP tendencies: - Deliberately seek out perspectives of distant others - Question judgments when we notice strong benefits to ourselves - Use abstract reasoning to counter motivated concrete reasoning - Engage in regular moral reflection and self-examination

Implication 3: Create Moral Communities

Since individual judgment is limited, we should participate in moral communities where diverse perspectives can challenge our biases.

Implication 4: Accept Moral Imperfection

Recognizing CMAP as a deep feature of human psychology means accepting that we will never achieve perfect moral consistency. The goal is not moral perfection but continuous improvement through reflective correction.

8. Conclusion : Toward a Realistic Yet Aspirational Moral Philosophy

This investigation of Contextual Moral Valuation and the Contextual Moral Alignment Paradox has revealed a complex picture of human moral judgment. We are neither the impartial rational calculators that traditional moral philosophy often assumes nor mere creatures of arbitrary prejudice and self-interest. Instead, we are *situated reasoners* agents whose moral judgments emerge from a particular position in social and evaluative space, shaped by cognitive architecture that reflects both our evolutionary history and our capacity for reflective correction.

8.1 Main Conclusions

Descriptive Conclusion: Human moral judgment integrates multiple evaluative dimensions (reward, threat, bias, distance) in systematic ways that produce context-dependent evaluation. This is not occasional inconsistency but a structural feature of moral cognition.

Normative Conclusion: Some forms of context-sensitivity are rationally defensible (sensitivity to epistemic access, special obligations, value pluralism), but much of CMAP constitutes moral failure (self-serving bias, arbitrary distance

effects, motivated reasoning). The challenge is distinguishing legitimate from illegitimate context-sensitivity.

Metaethical Conclusion: CMAP is compatible with moral objectivism if we recognize that moral judgment is necessarily situated. Moral truth can be objective even if moral judgment is perspectival, bounded, and context-sensitive. Progress involves developing capacities for reflective correction while acknowledging the limits of human moral cognition.

Practical Conclusion: Understanding CMAP has important implications for moral education, political reform, and individual moral life. Rather than demanding impossible impartiality, we should develop realistic strategies for recognizing and mitigating systematic biases.

8.2 A Realistic Yet Aspirational Framework

Traditional moral philosophy often presents an idealized picture of moral agents: fully informed, perfectly rational, capable of complete impartiality. This picture inspires but also misleads. When actual humans fail to meet this ideal (as they always do), we face a dilemma: dismiss moral philosophy as irrelevant to actual human life, or dismiss actual humans as hopelessly irrational.

The CMV framework suggests a third path: *realistic moral philosophy* that takes seriously both: 1. The cognitive and motivational architecture that constrains human moral judgment 2. Our capacity for reflective correction and moral improvement

This framework is :

Realistic because it acknowledges: - We cannot eliminate all context-sensitivity from moral judgment - Perfect impartiality is psychologically impossible for relational, embodied agents - Self-interest and psychological distance will always influence moral evaluation - Moral judgment operates under cognitive constraints requiring heuristics

Aspirational because it maintains : - We can distinguish better from worse forms of context-sensitivity - We can develop capacities for recognizing and correcting biases - Moral progress is possible through reflective equilibrium and institutional design - Moral ideals remain valuable even when perfect compliance is unattainable

8.3 Open Questions and Future Directions

This investigation opens several questions for future philosophical work :

Normative Questions: - Can we develop more precise criteria for distinguishing defensible from indefensible context-sensitivity? - What institutional designs best correct for CMAP tendencies? - How should moral philosophy balance ideal theory with realistic assessment of human capacities?

Metaethical Questions: - What does situated rationality imply for moral realism vs. anti-realism? - Can we develop a metaethics that accommodates both objective moral truth and perspectival moral judgment? - What is the relationship between moral motivation and moral judgment if they can come apart (as CMAP suggests)?

Applied Ethics Questions: - How should understanding CMAP inform policy in areas like international aid, climate change, corporate ethics, and criminal justice? - What obligations do we have to those harmed by our distant benefactors? - How can we design AI systems that avoid CMAP-like biases while being useful to humans?

8.4 Final Reflection

The phenomenon of selective moral alignment simultaneously condemning and sympathizing with the same agent initially appears as simple hypocrisy or moral failure. Yet deeper investigation reveals it as a window into the architecture of human moral cognition. This architecture is neither randomly inconsistent nor perfectly rational, but rather a complex system that reflects our evolutionary history, our cognitive limitations, our social embeddedness, and our capacity for reflective thought. Understanding this architecture does not excuse moral failures, but it does help us understand why such failures are systematic rather than random, why moral progress is difficult, and what strategies might be effective for improvement. It suggests that moral philosophy should be not merely prescriptive (telling us what we ought to do) but also diagnostic (helping us understand why we fail to do it) and therapeutic (providing realistic strategies for improvement). The Contextual Moral Alignment Paradox, far from being an embarrassing anomaly to be explained away, may be one of the most important phenomena for moral philosophy to address. For it reveals moral judgment not as the simple application of clear principles by detached reason, but as a complex process of integration, balancing, and navigation under conditions of uncertainty, limitation, and situatedness which is to say, under the conditions of actual human life.

References

- [1] Aristotle. (350 BCE). *Nicomachean Ethics*.
- [2] Bandura, A. (1999). Moral disengagement in the perpetration of inhumanities. *Personality and Social Psychology Review*, 3(3), 193-209.
- [3] Baron, J., & Spranca, M. (1997). Protected values. *Organizational Behavior and Human Decision Processes*, 70(1), 1-16.
- [4] Bloom, P. (2013). *Just Babies: The Origins of Good and Evil*. New York: Crown Publishers.

- [5] Cushman, F., Young, L., & Hauser, M. (2006). The role of conscious reasoning and intuition in moral judgment: Testing three principles of harm. *Psychological Science*, 17(12), 1082-1089.
- [6] Dancy, J. (2004). *Ethics Without Principles*. Oxford: Oxford University Press.
- [7] Darley, J. M., & Batson, C. D. (1973). From Jerusalem to Jericho: A study of situational and dispositional variables in helping behavior. *Journal of Personality and Social Psychology*, 27(1), 100-108.
- [8] Foot, P. (1967). The problem of abortion and the doctrine of double effect. *Oxford Review*, 5, 5-15.
- [9] Greene, J. D. (2013). *Moral Tribes: Emotion, Reason, and the Gap Between Us and Them*. New York: Penguin Press.
- [10] Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron*, 44(2), 389-400.
- [11] Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, 108(4), 814-834.
- [12] Haidt, J. (2012). *The Righteous Mind: Why Good People Are Divided by Politics and Religion*. New York: Vintage.
- [13] Kahane, G., Everett, J. A., Earp, B. D., Farias, M., & Savulescu, J. (2015). 'Utilitarian' judgments in sacrificial moral dilemmas do not reflect impartial concern for the greater good. *Cognition*, 134, 193-209.
- [14] Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2), 263-291.
- [15] Kant, I. (1785). *Groundwork of the Metaphysics of Morals*.
- [16] Kohlberg, L. (1969). Stage and sequence: The cognitive-developmental approach to socialization. In D. A. Goslin (Ed.), *Handbook of Socialization Theory and Research* (pp. 347-480). Chicago: Rand McNally.
- [17] Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 108(3), 480-498.
- [18] MacIntyre, A. (1981). *After Virtue*. Notre Dame: University of Notre Dame Press.
- [19] Mill, J. S. (1861). *Utilitarianism*. London: Parker, Son, and Bourn.
- [20] Nagel, T. (1986). *The View from Nowhere*. Oxford: Oxford University Press.
- [21] Nussbaum, M. C. (2001). *Upheavals of Thought: The Intelligence of Emotions*. Cambridge: Cambridge University Press.

- [22] Prinz, J. J. (2007). *The Emotional Construction of Morals*. Oxford: Oxford University Press.
- [23] Railton, P. (1984). Alienation, consequentialism, and the demands of morality. *Philosophy & Public Affairs*, 13(2), 134-171.
- [24] Rawls, J. (1971). *A Theory of Justice*. Cambridge, MA: Harvard University Press.
- [25] Scheffler, S. (1982). *The Rejection of Consequentialism*. Oxford: Oxford University Press.
- [26] Singer, P. (1972). Famine, affluence, and morality. *Philosophy & Public Affairs*, 1(3), 229-243.
- [27] Trope, Y., & Liberman, N. (2010). Construal-level theory of psychological distance. *Psychological Review*, 117(2), 440-463.
- [28] Williams, B. (1981). *Moral Luck*. Cambridge: Cambridge University Press.