
Moral Violations Are Made, Not Found

Evidence for Motivated Violation Construction

Emile Boullineau

Independent Researcher, UK

emile.boullineau@gmail.com

April 2026

Word count: ~3,800

Keywords *moral judgment • motivated reasoning • moral cognition • violation construction • moralization • social evaluation*

ABSTRACT

Moral psychology treats moral violations as features of behaviour that observers detect and then evaluate. This paper challenges this assumption. We introduce Motivated Violation Construction (MVC), the process whereby observers categorise ambiguous behaviour as morally wrong in order to justify predetermined negative evaluations.

Factors such as threat, dislike, envy, or group allegiance can motivate an observer to deem certain behaviour as morally wrong. This increases the likelihood that ambiguous behaviour is classified as a violation rather than simply judged more harshly. MVC is therefore activated at the point where behaviour first enters the moral domain. A stage before evaluation begins.

MVC is distinct from motivated reasoning, which shapes how people judge behaviour they already see as wrong. It is also distinguished from moralization, which concerns gradual shifts in what practices a society treats as moral issues. MVC, instead, is about how specific behaviour, expressed by a specific person, becomes classified as wrong.

Four predictions are advanced. Motivation increases the frequency with which identical behaviour is classified as a violation. When the behaviour is genuinely ambiguous, this effect is strongest. It weakens where moral consensus is clear. Differences in classification track the observer's motives rather than features of the behaviour itself. Critically, the classification of behaviour as wrong is driven more by motivation than by how severely it is judged once classified.

If proven, moral violations are not simply features of behaviour; they are partly determined by the observer. The implication is that they arise earlier than assumed by current models. Not in how behaviour is judged, but in whether it is seen as a violation at all.

1. THE DETECTION ASSUMPTION

Existing research on moral judgment typically treats moral violations as identified prior to evaluation. Consider the following cases. A person cuts in front of you in a queue. A colleague takes credit for work supported by others. A politician cancels a public commitment. In each of these cases, moral psychology assumes a common sequence: the observer perceives the behaviour, classifies it as a moral violation, and evaluates the actor accordingly. The violation is treated as a stable input across observers, something detected from the behaviour rather than something the observer brings to it.

This assumption is structural, not incidental. Attribution theory (Heider, 1958; Weiner, 1985) begins after the violation has already been identified, asking how people assign causes to recognised transgressions. Models of moral blame (Alicke, 2000; Malle *et al.*, 2014) specify how evaluators move from a perceived harmful event to a judgment of responsibility, but take the harmful event as given. Moral foundations theory (Graham *et al.*, 2013) maps the domains in which people make moral evaluations without asking whether the evaluative inputs are stable across observers. In every case, the moral violation functions as the independent variable: it is the stimulus that triggers judgment.

We call this the **detection assumption**: the implicit premise that moral violations exist in behaviour and are identified through a process that is fundamentally perceptual. Under this assumption, disagreements about whether a violation occurred reflect differences in moral standards, attention, or values. Current empirical literature does not address the possibility that the violation was constructed by the observer. Serving an underlying psychological need.

The detection assumption is reinforced by the dominant experimental paradigm in moral psychology. In vignette-based studies, the methodological backbone of the field, participants receive behaviours pre-classified as violations by the researcher. The participant evaluates a violation that has already been constructed for them. The paradigm does not permit the participant to decide whether the behaviour constitutes a violation at all. The result is a literature highly informative about how people process moral violations but largely silent about how those violations come into being.

2. MOTIVATED VIOLATION CONSTRUCTION

Every day life shows us that we react differently to certain behaviour depending on who the subject is. Consider watching someone you dislike arrive late to a meeting. The reaction is immediate: disrespectful, dismissive of everyone's time. But when a friend does the same thing, the same lateness becomes evidence of a busy schedule. Bluntness from a rival registers as rude, but when it's from a friend, it's evaluated as being refreshingly honest. The behaviour has not changed, but the moral label applied to it has.

This disconnect is not unusual and it's often observed in everyday life. Yet it causes a problem that has not been adequately addressed by moral psychology. If the same behaviour is classified as morally wrong or morally neutral depending on who performs it, and if this classification takes place before anything resembling formal moral reasoning has begun, then the violation cannot be a property of the behaviour. The observer's prior orientation toward the person is doing the classificatory work.

We introduce Motivated Violation Construction (or MVC) as our theoretical account and explanation of how this works. MVC proposes that an observer is more likely to classify ambiguous behaviour as a moral violation when a pre-existing negative motivation exists toward the target. This includes a perceived threat, envy, competition or intergroup bias. Therefore, under this mode, not all behaviour is more likely a violation. Ambiguous behaviour is.

Why does this happen? One plausible mechanism is based on justificatory need (cf. Kunda, 1990; Bastardi *et al.*, 2011). The observer already dislikes the target. But simple dislike without a moral basis produces dissonance. This is because most people want to regard themselves as fair-minded evaluators, not as people who simply dislike others without cause. To close that gap, the observer's cognition recruits moral classification itself as justification. An ambiguous act gets reclassified as an offence. And here is what makes MVC especially difficult to dislodge: the observer does not experience the constructed violation as constructed. It registers as genuine moral perception. The phenomenological invisibility of the process, the fact that the violation feels seen rather than made, means that exposing the underlying motivation rarely produces correction.

As far as the observer is concerned, they perceived what was there all along.

For acts such as theft, physical violence, and outright deception, the moral signal is loud enough that motivation barely changes the classification. Existing models of moral judgment handle these adequately and the intention is not for MVC to replace them. MVC operates in a different domain. The broad and poorly mapped space of everyday acts whose moral valence stays ambiguous to interpretation. Acts such as arriving late or speaking too bluntly. Withholding information or prioritising oneself are further examples of acts that can be interpreted as being a violation or as something acceptable. MVC argues that motivation is what moderates the framing toward one interpretation or the other. To be precise about the boundaries of this claim: observers do not fabricate behaviours or invent events. They adjust the threshold at which real, observed behaviours get slotted into the category of moral violation, often without any awareness that adjustment is occurring.

Two terms require clarification, because their confusion has obscured the process described by MVC. Classification is the threshold decision, the binary judgment that an act is wrong rather than permissible. This is where a behaviour enters the moral domain at all. Evaluation is what follows: the assessment of severity, the question of how wrong something is once it has been classified as a violation. MVC operates at the classification stage. Most existing models of moral judgment focus almost exclusively on evaluation. Because the two stages have not been clearly

pulled apart in the literature, the assumption that moral violations are simply detected, perceived like objects in a visual field, has gone largely unchallenged.

MVC describes a conditional process, not a personality trait. Any observer can construct violations when motivation toward a target is strong enough and the behaviour in question leaves room for more than one moral reading. Some people may do this more often or more readily than others, and that variation is worth studying (see Discussion). But the framework developed here is about the mechanism itself, not about who is more or less prone to deploying it.

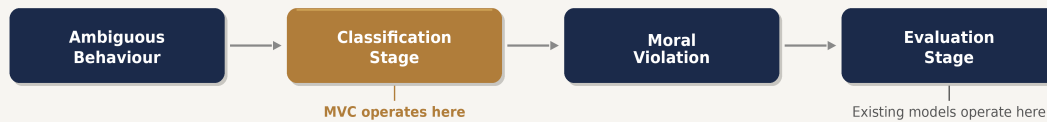


Figure 1. The locus of Motivated Violation Construction in the moral judgment processing chain. MVC operates at the classification stage (copper), where behaviour is first designated as a moral violation. Existing models begin at the evaluation stage.

3. DISTINGUISHING MVC FROM ADJACENT ACCOUNTS

Motivated Reasoning

Research on motivated reasoning (Kunda, 1990; Ditto & Lopez, 1992) is relevant here. It explains that people tend to arrive at conclusions by selectively searching for, interpreting, and weighing evidence. MVC takes this further and identifies a process that occurs earlier in the evaluative process. Conclusions are not only biased once a moral event has been recognised, but what counts as a moral event can itself be shaped by motivation in the first place. The object of judgment is not simply taken as given and can be constructed in ways that support the observer's prior evaluative stance. Motivated reasoning changes how a violation that is already recognised is subsequently evaluated. MVC, however, proposes that the violation itself may be constructed before evaluation begins.

Attribution Bias

Attributional models (Alicke, 2000) account for how an observer assigns disproportionate blame to disliked or threatening targets. What these models take for granted is that a blameworthy event has already been identified. The attribution process starts after violation detection. However, MVC operates before attribution happens. Attribution makes a detected violation seem worse, whereas MVC determines whether it gets detected in the first place.

Moralization

Research on moralization (Rozin, 1999; Brandt *et al.*, 2015) documents how preferences and practices migrate into the moral domain over time. Smoking acquires moral weight. Dietary choices become matters of right and wrong. MVC differs in its scope, mechanism, and temporal grain. It operates at the level of specific behavioural episodes, not social practices. It is driven by interpersonal motivation (dislike, envy, competition) rather than cultural or ideological change. It involves construction from individual acts in real time, not the gradual reclassification of entire categories. A person who moralizes meat-eating has undergone a value shift. A person who constructs a violation from a rival's ambiguous email has done something different entirely. This is motivated moral construction at the level of a single behavioural instance.

Harm Construction

The Theory of Dyadic Morality (Gray *et al.*, 2012; Schein & Gray, 2018) argues that moral judgment is fundamentally about perceived harm between an intentional agent and a suffering patient. What counts as harm is itself inferred rather than given. This account shares an important characteristic with MVC. Both constructs reject the idea that moral inputs are fixed properties of events waiting to be detected, but the two frameworks target different stages. Dyadic morality is concerned with the content of moral perception (specifically, whether an observer perceives harm in a given event), whereas MVC concerns the threshold (the point at which an observer classifies a behaviour as a moral violation at all), and the motivational conditions under which that threshold shifts.

These two things can come apart. An observer can perceive diffuse harm without classifying the behaviour as a violation (harm without moral classification). An observer can classify a behaviour as a violation under motivation without perceiving concrete harm (moral classification without harm). These accounts are not competing but rather complementary and operating at adjacent but distinct stages of the moral processing chain.

Motivated Moral Perception

Research on motivated moral perception (Gantman & Van Bavel, 2014, 2015) shows that moral stimuli capture visual attention and reach conscious awareness faster than non-moral stimuli. This perceptual advantage is in itself motivationally modulated. When justice motives are satiated, the advantage shrinks. The work establishes something that matters for MVC. Motivation shapes moral perception at the earliest stages of processing.

MVC and motivated moral perception diverge regarding assumptions about the input. Motivated moral perception accelerates the detection of stimuli that already carry moral weight. The signal is present in the environment and motivation acts as a perceptual filter that lowers the threshold for noticing it. MVC applies where no objective moral signal exists. The behaviour is

ambiguous, the moral content is not pre-existing, and motivation drives the observer to construct a violation from interpretive flexibility rather than to detect one from perceptual salience. In short: motivated moral perception assumes the presence of a moral signal. MVC specifies the conditions under which that signal is absent and constructed.

Table 1. Differentiating MVC from adjacent constructs.

Construct	What It Explains	Stage of Operation	Key Distinction from MVC
MVC	Observers construct violations from ambiguous behaviour to justify negative evaluations	<i>Pre-classification (upstream)</i>	—
Motivated Reasoning	Biased evidence search toward preferred conclusions	<i>Post-classification (conclusion-level)</i>	Biases the answer; MVC manufactures the question
Attribution Bias	Disproportionate blame to disliked targets	<i>Post-violation detection</i>	Presupposes a violation already identified; MVC creates it
Moralization	Preferences migrate into the moral domain over time	<i>Cultural / ideological (gradual)</i>	Domain-level shift in social norms; MVC is trial-level, motivation-driven
Harm Construction (Dyadic Morality)	Observers infer harm between agent and patient	<i>Content of moral perception</i>	Concerns <i>what</i> is harmful; MVC concerns <i>whether</i> the moral threshold is crossed
Motivated Moral Perception	Moral stimuli reach awareness faster; motivation modulates priority	<i>Perceptual detection (early)</i>	Detects pre-existing moral signals; MVC constructs signals where none exist

4. TESTABLE PREDICTIONS

MVC generates four core predictions. Each can be tested using standard experimental designs with between-subjects manipulation of motivation and within-subjects manipulation of behavioural ambiguity.

P1. Motivation increases violation labelling. Holding behaviour constant, the probability that behaviour is classified as morally wrong increases when the observer's motivation to evaluate it negatively is elevated. This represents identical behaviour that is subject to different moral classification. The key driver is the observer's motivational state and not the characteristics of the behaviour observed. This can be empirically tested using standard behavioural paradigms that manipulate target valence. e.g. by presenting brief biographical profiles with liked vs. disliked

indicators.

P2. The effect is strongest under ambiguity. MVC operates where behaviour is generally considered morally ambiguous. For behaviours with a widely accepted moral status, such as theft by a colleague (clearly wrong) or declining to attend a lunch event (clearly benign), the behavioural signals constrain classification and reduce the influence of motivation. The MVC effect should be most evident in the ambiguous zone: a manager reassigning a struggling employee's project without discussion, a friend cancelling plans at short notice, a co-worker forwarding a private message to a third party. These are acts whose moral status depends on how they are read. The prediction is that motivation determines the reading. The explicit failure condition that separates MVC from a general negativity bias is that, when behaviour is high-consensus (virtually all observers agree on its moral status), MVC should produce null or negligible effects. Pilot norming of behavioural stimuli across a non-motivated control sample would establish the necessary baseline classification rates.

P3. Constructed violations show a distinctive signature. Violations produced through MVC should show less agreement between observers and greater variation in how behaviour is classified than clear-cut transgressions. However, this variability will not be random. It should reflect differences in observers' motivations. Observers who share a similar stance toward the target are more likely to agree in classifying the behaviour as wrong, whereas those with different motivations are more likely to diverge. The key indicator is not variability alone, but variability that systematically follows motivation. This prediction can be tested by comparing agreement in moral classification between observers who share similar motivations and those who do not.

P4 (Critical Test). Classification is affected more than severity. This prediction most clearly distinguishes MVC from generalised evaluative bias and serves as the critical test of the framework. Motivation is expected to shift whether a behaviour is classified as wrong more than how severely it is judged once classified.

Practically, the influence of motivation on whether behaviour is seen as a violation should be stronger than its influence on how bad that behaviour is judged to be, given that it is already seen as wrong. If motivation had an equal effect on both, making everything seem worse, the effect would be better characterised as a general negativity bias. This would not require a new construct.

The key diagnostic feature is this asymmetry. It indicates that motivation operates at the point where behaviour is classified as a violation, rather than at the stage where its severity is assessed. This distinction can be tested using models that separate the likelihood of classifying behaviour as wrong from the severity of judgment once that classification is made.

Table 2. Summary of MVC predictions with diagnostic criteria and failure conditions.

Prediction	Core Claim	Diagnostic Criterion	Failure Condition
P1	Motivation increases violation labelling for identical behaviour	Higher classification rate for motivated vs. non-motivated observers	No difference in classification rates across motivation conditions
P2	MVC effect concentrated in ambiguous behaviours	Motivation × ambiguity interaction; null effect at high consensus	Equal effect across all ambiguity levels (= general negativity bias)
P3	Constructed violations show lower inter-rater consensus	Variability covaries with motivation (ICC across matched/mismatched pairs)	Variability is random and unrelated to motivation
P4 (Critical)	Motivation affects classification more than severity	Larger effect on binary wrong/not-wrong than conditional severity; hurdle model dissociation	Equal effect on classification and severity (= generalised evaluative bias)

5. DISCUSSION

Theoretical Implications

MVC's core premise is that a moral violation, a concept that is central to moral judgment research, is not a stable factor in the evaluative process.

For a substantial class of everyday behaviours, it can vary and be constructed by the observer. It carries motivational information about the observer as well as behavioural information about the target. What is usually treated as the independent variable in moral judgment research, the presence of a violation, is partially determined by the observer. This has methodological consequences. Vignette-based studies that present pre-defined violations to participants may be studying only the second half of the moral judgment process while bypassing the stage at which the most consequential variability occurs.

For motivated reasoning, MVC extends the account one stage earlier. People shape the starting point on which their conclusions depend, and do not simply reason their way to the conclusions they prefer. MVC shows how evaluation shapes perception. It reverses the usual assumption that perception comes first and evaluation comes after. We judge the person first and then decide that what they did was morally wrong.

Practical Implications

Political polarisation. MVC offers a mechanism for the well-documented observation that partisans perceive moral violations in the behaviour of political opponents that non-partisans do not see (Ditto *et al.*, 2019). Under MVC, the partisan is not judging the same violations more harshly. Rather, they are constructing violations from ambiguous political behaviour that a non-motivated observer would classify as legitimate disagreement. This constructive process may explain why factual corrections often fail to reduce partisan outrage: the outrage is grounded not in factual error but in motivationally constructed moral inputs experienced as genuine perception.

Interpersonal conflict. In close relationships, MVC predicts that accumulated negative motivation will progressively lower the threshold at which a partner's behaviour is classified as morally wrong. Acts once tolerated gradually enter the moral domain. The partner experiences this as inexplicable moral scrutiny; the motivated observer experiences it as a growing pattern of genuine transgressions. Both experiences are subjectively accurate, which is precisely what makes the conflict intractable.

Online moral discourse. Digital environments tend to strip away context while amplifying group identity and emotional reactions. People are often exposed to a high volume of behaviour that can be interpreted in different ways. These conditions make it easier for observers to see violations where others may not. MVC helps explain why moral outrage can escalate so quickly online, not simply as a breakdown of civility, but as a predictable result of motivated observers interpreting ambiguous behaviour in ways that fit their prior stance.

Relation to Broader Research on Evaluative Rigidity

MVC forms part of a broader research programme focusing on evaluative rigidity. The programme focuses on the psychological mechanisms behind how humans form evaluations. The framework examines situations where contextual information fails to change judgments even when context is fully acknowledged and understood, and introduces the evaluative gating model. This is a regulatory stage where contextual information either moderates the evaluation or does not. MVC focuses on a stage that occurs earlier. The stage before information reaches the evaluative gate. If motivation has already shaped the moral input, then the evaluative gate is operating on an already moralized input. In this case, two processes may combine. An observer may first notice a violation that a more flexible evaluator would not have noticed. The observer may fail to revise that judgment when new or mitigating contextual information becomes available. This effect remains a hypothesis rather than an established finding. However, the present paper does not attempt to specify how the gate operates. MVC instead focuses on how moral inputs come to be formed in the first place.

Limitations

MVC applies specifically to the ambiguous zone. For clear-cut and non-ambiguous transgressions, MVC makes no competing claim. The MVC framework does not specify whether violation construction is automatic or deliberative. This temporal question therefore, represents a priority for future work. At the mechanistic level, MVC is described in functional rather than neuropsychological terms. Future work may benefit from grounding violation construction in specific affective-cognitive architectures, such as Affective Realism (Barrett, 2017), which offers a promising account of how internal somatic states are projected onto external stimuli and experienced as objective features of the environment. The proposed predictions here are derived from existing evidence on motivated reasoning and moralization, along with first principles. Direct empirical tests are needed to confirm the account. Tests designed to permit participants to classify behaviour freely, rather than evaluating pre-constructed violations.

6. CONCLUSION

Moral psychology has largely built its theories on a foundation where violations are treated as given. MVC proposes that this foundation is incomplete. For the wide class of behaviours whose moral status is ambiguous, the violation is not found in the behaviour but is constructed by the observer. The construction is driven by motivation. Shaped by the observer's pre-existing stance toward the target, and experienced as genuine moral perception rather than motivated fabrication. If this account here is correct, then moral disagreement is not only about values, principles, or facts, it is also about what counts as a violation in the first place.

DECLARATIONS

Affiliations and Funding

This research is independent and is not affiliated with, funded by, or conducted on behalf of any employer, institution, or third party. The author has no institutional affiliation relevant to this work.

Use of AI-Assisted Tools

The author used AI-assisted tools (Claude, Anthropic) during the preparation of this manuscript for drafting assistance, editorial refinement, and formatting. The author reviewed, critically evaluated, and takes full responsibility for the theoretical arguments, conceptual claims, and scientific content presented in this paper.

Conflicts of Interest

The author declares no conflicts of interest.

REFERENCES

- Alicke, M. D. (2000). Culpable control and the psychology of blame. *Psychological Bulletin*, 126(4), 556–574.
- Barrett, L. F. (2017). *How emotions are made: The secret life of the brain*. Houghton Mifflin Harcourt.
- Bastardi, A., Uhlmann, E. L., & Ross, L. (2011). Wishful thinking: Belief, desire, and the motivated evaluation of scientific evidence. *Psychological Science*, 22(6), 731–738.
- Brandt, M. J., Wisneski, D. C., & Skitka, L. J. (2015). Moralization and the 2012 U.S. presidential election campaign. *Journal of Social and Political Psychology*, 3(2), 211–237.
- Ditto, P. H., & Lopez, D. F. (1992). Motivated skepticism: Use of differential decision criteria for preferred and nonpreferred conclusions. *Journal of Personality and Social Psychology*, 63(4), 568–584.
- Ditto, P. H., Liu, B. S., Clark, C. J., Wojcik, S. P., Chen, E. E., Grady, R. H., ... & Zinger, J. F. (2019). At least bias is bipartisan: A meta-analytic comparison of partisan bias in liberals and conservatives. *Perspectives on Psychological Science*, 14(2), 273–291.
- Gantman, A. P., & Van Bavel, J. J. (2014). The moral pop-out effect: Enhanced perceptual awareness of morally relevant stimuli. *Cognition*, 132(1), 22–29.
- Gantman, A. P., & Van Bavel, J. J. (2015). Moral perception. *Trends in Cognitive Sciences*, 19(11), 631–633.
- Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., & Ditto, P. H. (2013). Moral foundations theory: The pragmatic validity of moral pluralism. *Advances in Experimental Social Psychology*, 47, 55–130.
- Gray, K., Waytz, A., & Young, L. (2012). The moral dyad: A fundamental template unifying moral judgment. *Psychological Inquiry*, 23(2), 206–215.
- Heider, F. (1958). *The psychology of interpersonal relations*. Wiley.
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 108(3), 480–498.
- Malle, B. F., Guglielmo, S., & Monroe, A. E. (2014). A theory of blame. *Psychological Inquiry*, 25(2), 147–186.
- Rozin, P. (1999). The process of moralization. *Psychological Science*, 10(3), 218–221.
- Schein, C., & Gray, K. (2018). The theory of dyadic morality: Reinventing moral judgment by redefining harm. *Personality and Social Psychology Review*, 22(1), 32–70.
- Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological Review*, 92(4), 548–573.