Fallible but improvable: Repeated evaluation reduces moral inconsistency

Abstract

Extensive empirical research has documented inconsistencies in people's initial moral case judgments, which often result from order or framing effects and various cognitive biases. How people address such inconsistencies when given an opportunity to reconsider their case judgments has been less explored. Here, we examine whether people reduce inconsistency in their judgments on morally equivalent pairs of cases when given an opportunity to respond to cases again and in which of three settings they are most likely to do so: (i) repeated separate evaluation; (ii) joint evaluation; and (iii) joint evaluation preceded by an explicit prompt to assess their own consistency.

Using two exploratory studies – Pilot (N = 148) and Study 1 (N = 451) and a confirmatory online experiment Study 2 (N = 1,313), we find that participants tend to change their initial moral judgments to make them more consistent and that they are more likely to reduce their inconsistency under joint evaluation, when cases are presented together on one screen than under repeated separate evaluation, when they are presented in sequence on separate screens. However, we did not find strong evidence that the addition of an explicit consistency prompt – showing participants their own initial judgments and asking them to assess if those judgments were mutually consistent – further reduces inconsistency. While we observed that people who openly recognised their own inconsistency had a stronger tendency to reduce it, open recognition of inconsistency was neither sufficient nor necessary for reduction in inconsistency.

Our findings show that while people's initial judgments about morally equivalent cases are often inconsistent, people have a propensity to reduce their initial inconsistency when given an opportunity to reconsider their responses, even within the space of a short online experiment, and across different settings. This aligns with accounts of moral psychology, such as the moral consistency reasoning model, that posit that people are motivated to "treat like cases alike". This opportunity is most effective when inconsistency is made easy to spot by presenting cases together. Our results also suggest that people's moral judgments should not be dismissed as unreliable only because their initial responses to moral cases are often flawed. Elicitation methods which include scope for reconsideration allow people themselves to improve their judgments. However, our empirical research also highlights that people's

pursuit of consistency is limited, as a sizable share of inconsistencies, even openly recognised ones, remained unresolved.

Key words: moral judgement, moral inconsistency, joint evaluation

Tables

Table 6

Simple Inconsistencies ("Inconsistent responses") in the First Round and How They Were Addressed in the Second Round across Treatment Groups in Study 2.

Responses	Full sample		T0		T1		T2	
	N	%	n	%	n	%	n	%
All responses	6565		2195		2195		2175	
Consistent responses	2519	38		39		40		37
Inconsistent responses	4046	62		61		60		63
Remained the same	1228	30		37		25		29
Increased inconsistency	279	7		7		7		5
Reduced inconsistency	2539	63		56		68		66
Fully resolved inconsistency	1862	46		38		52		51

Table 8

Reduction in Inconsistency Score (IS-Reduction) across Treatment Groups in Study 2

Variable	Full sample	T0	T1	T2
IS-R1	5.16	5.11	5.10	5.26
IS-R2	3.16	3.68	2.98	2.82
IS-Reduction	**36%**	**23%**	**38%**	**48%**

Note. Inconsistency score is an aggregate metric that captures both the frequency and the extent of inconsistency. First, the extent of inconsistency was calculated for each pair of cases as an absolute difference between the responses to two paired scenarios (given we used a 6-point Likert scale, the smallest difference was 0, i.e., fully consistent response, and the greatest difference was 5, i.e., fully inconsistent response). Second, the values for all five paired cases were added up to form the overall inconsistency score. The minimum value for the inconsistency score was 0, the maximum was 25, and it was measured separately for 'initial responses' in the first round (IS-R1) and for 'reconsidered

judgments' in the second round (IS-R2). IS-Reduction is the average reduction in inconsistency between the first and the second round.

Outcomes of the main pre-registered analyses:

1. Testing if there was a difference between T1 and T0: We compared the IS-Reduction in the joint evaluation (T1) group and the repeated separate evaluation (T0) group. In line with our expectations, we found that IS-Reduction was higher under joint evaluation in T1 (M = 0.380, SD = 0.585) than under repeated separate evaluation in T0 (M = 0.231, SD = 0.528). Using two-sample Wilcoxon rank-sum (Mann–Whitney) test, as pre-registered, we confirmed that the difference of 15 percentage points (MD = −0.149) was statistically significant (z = −5.54, p < .001, Cohen's d = −0.27, 95% CI [−0.40, −0.13], N = 861).
2. Testing if there was a difference between T1 and T0: We compared the IS-Reduction in the joint evaluation (T1) and the joint evaluation with an explicit consistency prompt (T2) group. While the reduction in inconsistency score was higher in T2 (48%) than in T1 (38%); however, the difference was not statistically significant. Based on the outcomes of the two-sample Wilcoxon rank-sum (Mann–Whitney) test, we were not able to reject the null hypothesis (full sample: MD = −0.098, z = −1.82, p = .069, N = 862).

Bibliography

Andow, J. (2016). Reliable but not home free? What framing effects mean for moral intuitions. *Philosophical Psychology*, *29*(6), 904-911.

Baron, J., Scott, S., Fincher, K., & Metz, S. E. (2015). Why does the Cognitive Reflection Test (sometimes) predict utilitarian moral judgment (and other things)?. *Journal of Applied Research in Memory and Cognition*, *4*(3), 265-284.

Bengson, J. (2013). Experimental attacks on intuitions and answers. *Philosophy and Phenomenological Research*, 86(3), 495–532. https://doi.org/10.1111/j.1933-1592.2012. 00578.x

Brannon, S. M., Sacchi, D. L. M., & Gawronski, B. (2017). (In)consistency in the eye of the beholder: The roles of warmth, competence, and valence in lay perceptions of inconsistency. *Journal of Experimental Social Psychology*, 70, 80–94.

Campbell, R., & Kumar, V. (2012). Moral reasoning on the ground. *Ethics,* 122(2), 273–312.

Chi, M. T., & Ohlsson, S. (2005). Complex declarative learning. In K. J. Holyoak & R. G. Morrison (Eds.), *Cambridge Handbook of Thinking and Reasoning* (pp. 371–399). Cambridge University Press.

Cialdini, R. B., Trost, M. R., & Newsom, J. T. (1995). Preference for consistency: The development of a valid measure and the discovery of surprising behavioral implications. *Journal of Personality and Social Psychology*, *69*(2), 318.

Cooper, J., & Fazio, R. H. (1984). A new look at dissonance theory. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 17, pp. 229–266).

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.

Daniels, N. (1979). Wide reflective equilibrium and theory acceptance in ethics. *The Journal of Philosophy*, *76*(5), 256-282.

De Quidt, J., Haushofer, J., & Roth, C. (2018). Measuring and bounding experimenter demand. *American Economic Review*, *108*(11), 3266-3302.

Demaree-Cotton, J. (2016). Do framing effects make moral intuitions unreliable?. *Philosophical Psychology,* 29(1), 1-22.

Doris, J., & Stich, S. (2005). As a matter of fact: Empirical perspectives on ethics. In F. Jackson and M. Smith (Eds.), *The Oxford handbook of contemporary analytic philosophy* (pp. 114-152). Oxford: Oxford University Press.

Dreisbach, S., & Guevara, D. (2019). The Asian disease problem and the ethical implications of prospect theory. *Noûs*, 53(3), 613-638.

Effron, D. A., & Helgason, B. A. (2023). Moral inconsistency. In B. Gawronski (Ed.), *Advances in experimental social psychology* (Vol. 67, pp. 1–72). Academic Press.

Eftedal, N. H., & Thomsen, L. (2021). Motivated moral judgments about freedom of speech are constrained by a need to maintain consistency. *Cognition*, *211*, 104623.

Epley, N., & Gilovich, T. (2016). The mechanics of motivated reasoning. *Journal of Economic Perspectives*, 30(3), 133–40.

Fagerlin, A., Zikmund-Fisher, B.J., Ubel, P.A., Jankovic, A., Derry, H.A., & Smith, D.M. (2007) Measuring numeracy without a math test: Development of the Subjective Numeracy Scale (SNS). *Medical Decision Making*, *27*, 672–680.

Festinger, L. (1957). *A theory of cognitive dissonance.* Stanford University Press.

Gawronski, B., & Brannon, S. M. (2019). What is cognitive consistency, and why does it matter? In E. Harmon-Jones (Ed.), *Cognitive dissonance: Reexamining a pivotal theory in psychology* (2nd ed., pp. 91–116). American Psychological Association.

Gino, F., Shu, L. L., & Bazerman, M. H. (2010). Nameless+ harmless= blameless: When seemingly irrelevant factors influence judgment of (un) ethical behavior. *Organizational Behavior and Human Decision Processes*, *111*(2), 93-101.

Gosling, S. D., Rentfrow, P. J., & Swann, W. B., Jr. (2003). A Very Brief Measure of the Big Five Personality Domains. *Journal of Research in Personality, 37*, 504–528.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, *108*(4), 814–834.

Harmon-Jones, E., Amodio, D. M., & Harmon-Jones, C. (2009). Action-based model of dissonance: A review, integration, and expansion of conceptions of cognitive conflict.

In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 41, pp. 119–166). Elsevier Academic Press.

Harmon-Jones, E., Harmon-Jones, C., & Levy, N. (2015). An action-based model of cognitive-dissonance processes. *Current Directions in Psychological Science*, 24(3), 184–189.

Holyoak, K. J., & Powell, D. (2016). Deontological coherence: A framework for commonsense moral reasoning. *Psychological Bulletin,* 142(11), 1179.

Horne, Z., & Livengood, J. (2017). Ordering effects, updating effects, and the specter of global skepticism. *Synthese*, *194*(4), 1189-1218.

Horne, Z., Powell, D., & Hummel, J. (2015). A single counterexample leads to moral belief revision. *Cognitive Science, 39*(8), 1950–1964.

Horne, Z., Powell, D., & Spino, J. (2013). Belief updating in moral dilemmas. *Review of Philosophy and Psychology*, *4*(4), 705-714.

Horne, Z., Rottman, J., & Lawrence, C. (2021). Can coherence-based interventions change dogged moral beliefs about meat-eating?. *Journal of Experimental Social Psychology*, *96*, 104160.

Hsee, C. K., Loewenstein, G. F., Blount, S., & Bazerman, M. H. (1999). Preference reversals between joint and separate evaluations of options: A review and theoretical analysis. *Psychological bulletin*, *125*(5), 576.

Kahan, D. M. (2016). The politically motivated reasoning paradigm. In R. A. Scott & S. M. Kosslyn (Eds.), *Emerging trends in the social and behavioral sciences*. Retrieved May 18, 2023, from https://onlinelibrary.wiley.com/doi/10.1002/9781118900772.etrds0417

Kahane, G. (2014). Intuitive and counterintuitive morality. In D. Jacobson & J. D'Arms (Eds.), *Moral psychology and human agency* (pp. 9–39*)*. Oxford University Press.

Kahneman, D., Knetsch, J. L., & Thaler, R. (1986). Fairness as a constraint on profit seeking: Entitlements in the market. *American Economic Review*, 76(4), 728-741.

Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Bahník, Š., Bernstein, M. J., … Nosek, B. A. (2014). Investigating Variation in Replicability: A "Many Labs" Replication Project. *Social Psychology, 45*(3), 142–152. https://doi.org/10.1027/1864-9335/a000178

Knobe, J. (2003). Intentional action and side effects in ordinary language. *Analysis*, *63*(3), 190-194.

Konrath, S., Meier, B. P., & Bushman, B. J. (2018). Development and validation of the single item trait empathy scale (SITES). *Journal of Research in Personality*, 73, 111–122.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 108(3), 480.

Lammers, J. (2012). Abstraction increases hypocrisy. *Journal of Experimental Social Psychology, 48*(2), 475-480.

Levy, N., Harmon-Jones, C., & Harmon-Jones, E. (2018). Dissonance and discomfort: Does a simple cognitive inconsistency evoke a negative affective state? *Motivation Science, 4*(2), 95–108. https://doi.org/10.1037/mot0000079

Lombrozo, T. (2009). The role of moral commitments in moral judgment. *Cognitive Science, 33*(2), 273-286.

Luptáková, V., Galizzi, M. M., Dario, K., & Voorhoeve, A. (2025). Do people better recognize inconsistencies in others' moral judgments than their own? If so, why? *Philosophical Psychology*, 1–33. https://doi.org/10.1080/09515089.2025.2558902

Meyer, M., Alfano, M., & de Bruin, B. (2024). The development and validation of the Epistemic Vice Scale. *Review of Philosophy and Psychology*, *15*, 355-382. https://doi.org/10.1007/s13164-021-00562-5

Meyers, E. A., Białek, M., Fugelsang, J. A., Koehler, D. J., & Friedman, O. (2019). Wronging past rights: The sunk cost bias distorts moral judgment. *Judgment and Decision Making*, *14*(6), 721–727.

Mummolo, J., & Peterson, E. (2019). Demand effects in survey experiments: An empirical assessment. *American Political Science Review*, *113*(2), 517-529.

Nadelhoffer, T., & Feltz, A. (2008). The actor–observer bias and moral intuitions: Adding fuel to Sinnott-Armstrong's fire. *Neuroethics*, *1*, 133-144.

Newby-Clark, I. R., McGregor, I., & Zanna, M. P. (2002). Thinking and caring about cognitive inconsistency: When and for whom does attitudinal ambivalence feel uncomfortable?. *Journal of Personality and Social Psychology*, *82*(2), 157-166.

Paharia, N., Kassam, K. S., Greene, J. D., & Bazerman, M. H. (2009). Dirty work, clean hands: The moral psychology of indirect agency. *Organizational Behavior and Human Decision Processes*, *109*(2), 134–141.

Paxton, J. M., & Greene, J. D. (2010). Moral reasoning: Hints and allegations. *Topics In Cognitive Science, 2*(3), 511-527.

Petrinovich, L., & O'Neill, P. (1996). Influence of wording and framing effects on moral intuitions. *Ethology and Sociobiology, 17*, 145–171.

Pizarro, D. A., & Bloom, P. (2003). The intelligence of the moral intuitions: A comment on Haidt (2001). *Psychological Review, 110*(1), 193-198.

Rai, T. S., & Holyoak, K. J. (2010). Moral principles or consumer preferences? Alternative framings of the trolley problem. *Cognitive Science*, *34*(2), 311–321.

Rawls, J. (1951). Outline of a decision procedure for ethics. *The Philosophical Review*, *60*(2), 177-197.

Rawls, J. (1999). *A theory of justice* (Revised ed.). Belknap Press. (Original work published 1971)

Rehren, P., & Sinnott-Armstrong, W. (2021). Moral framing effects within subjects. *Philosophical Psychology*, *34*(5), 611-636.

Rehren, P., & Sinnott-Armstrong, W. (2023). How stable are moral judgments?. *Review of Philosophy and Psychology*, *14*(4), 1377-1403.

Rini, R. A. (2015). How not to test for philosophical expertise. *Synthese,* 192, 431-452.

Ross, W.D. (1988). *The right and the good*. Clarendon Press.

Rugg, D. (1941). Experiments in wording questions: II. *Public Opinion Quarterly, 5*, 91–92.

Sauer, H. (2021). Against moral judgment. The empirical case for moral abolitionism. *Philosophical Explorations*, *24*(2), 137-154.

Scanlon, T. (1998). *What we owe to each other.* The Belknap Press of Harvard University Press.

Schwitzgebel, E., & Cushman, F. (2012). Expertise in moral reasoning? Order effects on moral judgment in professional philosophers and non-Philosophers. *Mind & Language, 27*(2), 135–153.

Schwitzgebel, E., & Cushman, F. (2015). Philosophers' biased judgments persist despite training, expertise and reflection. *Cognition*, *141*, 127-137.

Sinnott-Armstrong, W. (2008). Framing moral intuitions. In W. Sinnott-Armstrong (Ed.), *Moral psychology: The cognitive science of morality: Intuition and diversity* (Vol. 2, pp. 47–76). MIT Press.

Stanovich, K. E. (2018). Miserliness in human cognition: The interaction of detection, override and mindware. *Thinking & Reasoning, 24*(4), 423-444.

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences, 23*(5), 645–665.

Stanovich, K., & West, R. F. (1998). Individual differences in framing and conjunction effects. *Thinking & Reasoning,* 4(4), 289-317.

Sunstein, C. R., Kahneman, D., Schkade D., & Ritov, I. (2001). Predictably incoherent judgments. *John M. Olin Program in L. & Econ.* Working Paper No. 131.

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *Science*, 185(4157), 1124-1131.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211*(4481), 453–458.

Wiegmann, A., & Waldmann, M. R. (2014). Transfer effects between moral dilemmas: A causal model theory. *Cognition*, *131*(1), 28–43.

Zamzow, J. L., & Nichols, S. (2009). Variations in ethical intuitions. *Philosophical Issues*, 19, 368-388.

Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. *Experimental Economics*, *13*, 75-98.