Joint Action at the Roots of Ethical Cognition

Stephen A. Butterfill
< s.butterfill@warwick.ac.uk >

Thursday, 12th June 2025

Contents

1. Introduction

Here is an old, now quite familiar idea about ethics:

> 'The moral sentiments were designed for … hunter-gatherer villages and other, earlier, societies that are lost in the mists of prehistory. It is safe to say that these societies didn't have an elaborate judicial system and a large police force.' (Wright 1994, p. 374)[1]

This idea guides thinking about ethical improvement. To illustrate:

> 'We have an empirically confirmed theory about where our moral judgments come from [...] This amounts to the discovery that our moral beliefs are products of a process that is entirely independent of their truth, which forces the recognition that we have no grounds one way or the other for maintaining these beliefs.' (Joyce 2006, p. 211)[2]

But is the idea about evolution correct?

The arguments that Wright (1994) and Joyce (2006) offer depend on premises about how humans come have to moral intuitions and how they make ethical judgements. This area, moral psychology, has seen many new discoveries in the last decade.

Key support for the idea about evolution comes from discoveries associated with Greene et al's dual-process theory of moral cognition (Greene 2014b, 2017).

If we found that the dual-process theory of moral cognition is not well supported by evidence, this would remove a key support for the idea about intuitions or sentiments reflecting humans' evolutionary history.[3] …

---

[1] See also Wright (1994, p. 191): 'We live in cities and suburbs and watch TV and drink beer, all the while being pushed and pulled by feelings designed to propagate our genes in a small hunter-gatherer population.' This is a version of what call the Richerson & Boyd (2005, p. 150) 'the big-mistake hypothesis'.

[2] Others offer less radical conclusions but endorse the claim that the idea about evolution should guide our thinking about ethics. See, for example, Vlerick (2017, p. 238): 'An evolutionary perspective on our moral wiring … teaches us to regard the output of some of our evolved moral dispositions with a healthy dose of skepticism, given that the behavior they evolved to produce cannot always be expected to lead to good (moral) results (as determined by our reason-powered moral compass).'

[3] Boyd & Richerson (2005) argue that Wright, Joyce and others are wrong about the scale of prehistric social systems, and the identify a significant role for cultural evolution. They agree that 'Contemporary human societies differ drastically from the societies in which our social instincts evolved' (p. 264) and also hold that 'Innate principles furnish people

2. Mixed Evidence for a Dual-Process Theory of Ethical Cognition

2.1. Greene et al's Dual-Process Theory

Greene et al offer a dual-process theory of ethical cognition:

> 'this theory associates controlled cognition with utilitarian (or consequentialist) moral judgment aimed at promoting the "greater good" (Mill, 1861/1998) while associating automatic emotional responses with competing deontological judgments that are naturally justified in terms of rights or duties (Kant, 1785/1959).' (Greene 2015, p. 203)

The theory was developed in part to explain otherwise apparently anomalous responses to moral dilemmas. In particular, people have substantially different attitudes to killing one person in order to save several others depending on whether the killing involves pressing a switch (as in the Switch dilemma) or whether it involves dropping someone through a trapdoor into the path of great danger (as in the Footbridge dilemma).[4]

What is the explanation Greene et al's theory offers?

> 'this pattern of judgment [Switch—yes; Footbridge—no] reflects the outputs of distinct and (in some cases) competing neural systems [...] The more "personal" harmful action in the footbridge case, pushing the man off the footbridge, triggers a relatively strong negative emotional response, whereas the relatively impersonal harmful action in the switch case does not.' (Greene 2015, pp. 203—4)

2.2. Mixed Behavioural Evidence for This Dual-Process Theory

One prediction of the theory is that increasing time pressure should increase the influence of automatic emotional processes relative to the influence of controlled cognition, which in turn should make responses that are characteristically deontological more likely.

---

with basic predispositions, emotional capacities, and social dispositions that are implemented in practice through highly variable cultural institutions' (p. 264).

[4] See Greene (2015, p. 203): 'We developed this theory in response to a long-standing philosophical puzzle ... Why do people typically say "yes" to hitting the switch, but "no" to pushing?'

This prediction is supported by (Suter & Hertwig 2011), among others.[5] But Bago & De Neys (2019) consider what happens when subjects first make a moral judgement under time pressure and extraneous cognitive load and then, just after, make another moral judgement (in answer to the same question) with no time pressure and no extraneous cognitive load. They report:

> 'Our critical finding is that although there were some instances in which deliberate correction occurred, these were the exception rather than the rule. Across the studies, results consistently showed that in the vast majority of cases in which people opt for a [consequentialist] response after deliberation, the [consequentialist] response is already given in the initial phase' (Bago & De Neys 2019, p. 1794).

Rosas & Aguilar-Pardo (2020) find, conversely to what Greene et al's theory predicts, that subjects are less likely to give characteristically deontological responses under extreme time pressure.

The converse finding of Rosas & Aguilar-Pardo (2020) is not theoretically unmotivated—there are also some theoretical reasons for holding that automatic emotional processes should support characteristically utilitarian responses (Kurzban et al. 2012).

As there is a substantial body of neuropsychological evidence in favour of Greene et al's theory (reviewed in Greene 2014a), its defenders may be little moved by the mixed behavioural evidence. But there is a reason, not decisive but substantial, to expect mixed evidence more generally …

2.2.1. Methodological Challenge

The mixed pattern of evidence for and against Greene et al's theory might be explained by their choice of vignettes using trolley cases as stimuli. Waldmann et al. (2012, p. 288) offers a brief summary of some factors which have been considered to influence responses including:

- whether an agent is part of the danger (on the trolley) or a bystander;
- whether an action involves forceful contact with a victim;
- whether an action targets an object or the victim;
- how far the agent is from the victim;[6] and

[5] See also Trémolière & Bonnefon (2014) and Conway & Gawronski (2013) (who manipulated cognitive load).

[6] After this review was published, Nagel & Waldmann (2013) provided substantial evidence that distance may not be a factor influencing moral intuitions after all (the impression that it does was based on confounding distance with factors typically associated with distance such as group membership and efficacy of action).

- how the victim is described.

Other factors include whether there are irrelevant alternatives (Wiegmann et al. 2020); and order of presentation (Schwitzgebel & Cushman 2015).

They comment:

> 'A brief summary of the research of the past years is that it has been shown that almost all these confounding factors influence judgments, along with a number of others [...] it seems hopeless to look for the one and only explanation of moral intuitions in dilemmas. The research suggests that various moral and nonmoral factors interact in the generation of moral judgments about dilemmas' (Waldmann et al. 2012, pp. 288, 290).

For proponents of Greene et al's view, this might be taken as encouragement. Yes, the evidence is a bit mixed. But perhaps what appears to be evidence falsifying predictions of the view will turn out to be merely a consequence of extraneous, nonmoral factors influencing judgements.

Alternatively, Waldmann et al.'s observation could be taken to suggest that few if any of the studies relying on dilemmas presented in vignette form provide reliable evidence about moral factors since they do not adequately control for extraneous, nonmoral factors. As an illustration, Gawronski et al. (2017) note that aversion to killing (which would be characteristically deontological) needs to be separated from a preference for inaction. When considering only aversion to killing, time pressure appears to result in characteristically deontological responses, which would support Greene et al's theory (Conway & Gawronski 2013). But when aversion to killing and a preference for inaction are considered together, Gawronski et al. (2017) found evidence only that time pressure increases preferences for inaction.

While the combination of mixed behavioural evidence and methodological challenges associated with using dilemmas presented in vignettes does not provide a case for rejecting Greene et al's view, it does motivate considering fresh alternatives.

2.3. Suggestion

While we have not seen decisive evidence against it, we have seen enough to motivate seeking alternatives.

3. Norms-like Patterns

Does normative guidance require normative attitudes?

3.1. Normative Attitudes

Some researchers characterise norms in such a way that normative guidance requires normative attitudes by definition. For example:

> 'norms are characterized by general acceptance of particular normative principles within the group in question.' (Brennan et al. 2013, p. 94)

Bicchieri offers a theory that allows for normative guidance by habits. However, Bicchieri's envisaged role for habits is conditional on their functioning in ways that, on her account, amount to attitudes controlling behaviors (Bicchieri 2005, p. 6). As she puts it: normative behaviors must be under the control of attitudes in the less demanding sense that if the attitudes were to change, the behaviors would 'at once be overridden and abandoned' (Bicchieri 2005, 51).

Brennan et al. (2013, pp. 28ff) trace their idea back to Hart's distinction between rules and habits:

> 'How does a habit [mere pattern in behaviour] differ from a rule? ... A social rule has an 'internal' aspect ... there should be a critical reflective attitude to certain patterns of behaviour as a common standard ... this should display itself in criticism (including self-criticism), demands for conformity, and in acknowledgements ..., all of which find their characteristic expression in the normative terminology of 'ought" (Hart 1994, p. 55–7)

3.2. Normative Regularities and implicit Normative Behaviours

Other researchers characterise social norms in ways that do not involve attitudes at all. For example, Westra & Andrews (2022) define a 'normative regularity' as 'a socially maintained pattern of behavioral conformity within a community.' This is helpful because it suggests we can characterise normative phenomena in a way that leaves open for discovery questions about which mechanisms are responsible for them.

Heyes (2024) proposes a notion of 'implicit normative behaviours' which involve compliance and enforcement but, unlike their explicit counterparts, do not involve commentary. Her key insight is that such implicit normative behaviours can be a consequence of refinforcement learning and do not require normative attitudes.

Despite these virtues, neither approach seems well-suited to understanding normative guidance as it stands because there do not appear to be non-instrumental reasons to conform to normative regularities nor to maintain

implicit normative behaviours.

As first step to characterising a form of normative guidance require normative attitudes, consider the notion of a norm-like pattern.

3.3. Norm-Like Patterns

An act of rewarding, punishing or signaling compliance is *normatively catalytic* just if there is a pattern of behaviour which the act contributes to upholding.

A *norm-like pattern* is a pattern of behaviour which exists because > 1. there are normatively catalytic actions concerning this pattern; and > > 2. some or all of these actions have the collective goal of upholding this pattern.

Where there are norm-like patterns, we can ask which mechanisms are responsible for the normatively catalytic actions.

A special case of norm-like patterns are those where the normatively catalytic actions are driven by normative attitudes. These patterns count as norms in the sense Hart (1994) and others focus on.

But there are also norm-like patterns where the normatively catalytic actions are driven by more basic mechanisms. including reinforcement learning.

3.4. Collective Goals

The notion of a norm-like pattern requires that of a collective goal. What are collective goals?

A goal is an outcome to which one or more actions are directed. Someone might say, for example, 'the goal of our actions is to free Nelson Mandela.' Note that a goal is not an intention, nor any mental state of the agents. (At least, not usually.) The freedom of Nelson Mandela is not a mental state of those who ensured his freedom.

An outcome is a collective goal of two or more actions involving multiple agents if it is an outcome to which those actions are directed where this is not, or not only, a matter of each action being directed to the outcome.

Can you give sufficient conditions for there to be a collective goal? Yes!

> If there is a single outcome, G, such that
>
> 1. Our actions are coordinated; and
> 2. coordination of this type would normally increase the probability that G occurs.

> then there is an outcome to which our actions are directed where this is not, or not only, a matter of each action being directed to that outcome, i.e. our actions have a collective goal.

3.5. Background on Interspecies Norm-like Patterns

> 'Trophobiotic interactions involve the consumption of a food reward, often in return for protection from natural enemies. For ant-loving hemipterans, caterpillars, and most plants, these rewards almost invariably involve a sugary and/or nutrient-rich liquid, one that is collected by the foragers that patrol the area surrounding the resource ... Highly specialized ant-plants (myrmecophytes) offer additional food rewards and provide ants with a domicile.' (Ness et al. 2010, p. 99)

3.6. Minimal Normative Guidance

The notion of a norm-like pattern does **not**, by itself, involve anything normative. Here we face a dilemma:

> Insofar as a norm-like pattern is merely a result of natural selection, or of reinforcement learning, there is **no reason** why anyone should conform to the pattern. Insofar as a norm-like pattern is driven by normative attitudes, there is a clear sense in which normative guidance is involved, but this is exactly not what we are looking for.

How could upholding a pattern be a collective goal of our actions other than through natural selection, mere reinforcement learning, explicit intentions or normative attitudes?

Observation: violations of ethical norms can cause feelings such as bitterness, disgust or social pain, and such feelings can also influence ethically-relevant behaviours and evaluations.[7]

These effects, although significant and often replicable, are small (Landy & Goodwin 2015; Chapman 2018; Piazza et al. 2018; Giner-Sorolla et al. 2018), vary from person to person in ways we do not understand (e.g. Tracy et al. 2019), and are culturally mediated (e.g. Terrizzi et al. 2010).

Here is a conjecture:

[7] Chapman et al. (2009); Eskine et al. (2011) on bitterness; (Tracy et al. 2019; Vanaman & Chapman 2020; Chapman & Anderson 2013; Lai et al. 2014; Giner-Sorolla & Chapman 2017) on disgust; (Gawronski et al. 2018) on happiness; and MacDonald & Leary (2005) on social pain.

> Disgust, bitterness, social pain and other feelings function to enable us to create patterns of behaviour, which are thereby norm-like.

Where norm-like patterns are the result of feelings like these, and if these feelings really do exist in part because they enable us to create norm-like patterns, they they provide a form of normative guidance.

This is not the same form of normative guidance provided by attitudes (as when we all agree that we should follow a rule about not stealing each other's chocolate). But it is *a* form of normative guidance in this sense: where our coordination on what is bitter or disgusting or socially painful results in a norm-like pattern we impose on ourselves, it can be reasonable for us judge that we should conform to a norm-like pattern even without knowing why we should.

3.7. Metacognitive Feelings

Examples include the feeling of familiarity (Whittlesea & Williams 1998; Scott & Dienes 2008), of knowing (Koriat 2000), and of being the agent of an event (Haggard & Chambon 2012).

Koriat's theory:

> 'metacognitive feelings are mediated by the implicit application of nonanalytic heuristics [... which] operate below full consciousness, relying on a variety of cues [... and] affect metacognitive judgments by influencing subjective experience itself' (Koriat 2000, p. 158; see also Koriat 2007, pp. 313–5).

Metacognitive feelings typically arise from processes which monitor cognitive fluency, although the strength of feeling tends to be linked to how unexpected the fluency (or its lack) is. For example, the strongest feeling of familiarity is obtained from faces which are easy to process but difficult to identify explicitly (Whittlesea & Williams 1998).

As this suggests, metacognitive feelings involve interpretation. Further, it is possible to learn a novel interpretation for a metacognitive feeling. For example, Wan et al. (2008) trained participants to use familiarity in deciding whether a stimulus is from that grammar

4. Conclusion

4.1. Do all forms of normative guidance require attitudes?

Not if the conjecture that socially coordinated feelings of disgust, bitterness, or pain function to enable us to adopt norm-like patterns is correct.

4.2. Why does it feel wrong when we know it isn't?

Because (so my speculative proposal)

1. actual or anticipated violations of norm-like patterns create (i) metacognitive feelings of disfluency and (ii) anticipation of sanctions or of feelings of disgust, bitterness or pain,
2. where the anticipation of sanctions or feelings unconsciously biases me to interpret the feeling of disfluency as wrongness.

Note that on this view, feelings of disgust, bitterness or pain are only indirectly directly tied to the feeling of wrongness (via metacognitive disfluency).[8]

4.3. How do incidental feelings influence moral judgements?

Indirectly. They may play a role in establishing and maintaining norm-like patterns. And they could play a role in biasing us to interpret metacognitive feelings of disfluency associated with violations of norm-like patterns as feelings of wrongness.

4.4. What is the best computational description of fast ethical processes?

Wildly speculative conjecture: the processes likely to be evolutionarily ancient, to appear early in development and which influence adults' ethical intuitions involve coordination around norm-like patterns.

[8] As Piazza et al. (2018, p. 73) note, there is 'lack of covariance between felt disgust and wrongness judgments'. See also Giner-Sorolla et al. (2018, p. 264): 'ambient incidental disgust does not seem to have strong and consistent effects on moral judgment.'

4.5. Are evolutionarily ancient processes which appear early in development guided by principles adaptive in prehistoric environments?

Not if the conjecture that these processes involve socially coordinated feelings of disgust, bitterness, or pain which function to enable norm-like patterns is correct.

The processes do not encode principles. However they may appear to do so insofar as some things are most likely to provoke feelings of bitterness, disgust or pain.

4.6. Why is there a gap between material and ethical understanding?

Because (i) intuitions are recognized a constraint by many researchers in ethics, and (ii) one source of these intuitions is norm-like patterns which are not suited to theoretical generalization.

Glossary

characteristically deontological According to Greene, a judgement is *characteristically deontological* if it is one in 'favor of characteristically deontological conclusions (eg, "It's wrong despite the benefits")' (Greene 2007, p. 39). According to Gawronski et al. (2017, p. 365), 'a given judgment cannot be categorized as deontological without confirming its property of being sensitive to moral norms.' 4, 5

collective goal an outcome to which two or more agents' actions are directed where this is not, or not only, a matter of each action being directed to that outcome (Butterfill & Sinigaglia 2022). 8

Drop A dilemma; also known as *Footbridge*. A runaway trolley is about to run over and kill five people. You can hit a switch that will release the bottom of a footbridge and one person will fall onto the track. The trolley will hit this person, slow down, and not hit the five people further down the track. Is it okay to hit the switch? 13

dual-process theory Any theory concerning abilities in a particular domain on which those abilities involve two or more processes which are distinct in this sense: the conditions which influence whether one process occurs differ from the conditions which influence whether another occurs. 4

Footbridge A dilemma; also known as *Drop*. A runaway trolley is about to run over and kill five people. You can hit a switch that will release the bottom of a footbridge and one person will fall onto the track. The trolley will hit this person, slow down, and not hit the five people further down the track. Is it okay to hit the switch? 4

goal A *goal* of an action is an outcome to which it is directed. 8

outcome An outcome of an action is a possible or actual state of affairs. 8, 12

Switch A dilemma; also known as *Trolley*. A runaway trolley is about to run over and kill five people. You can hit a switch that will divert the trolley onto a different set of tracks where it will kill only one. Is it okay to hit the switch? 4

Transplant A dilemma. Five people are going to die but you can save them all by cutting up one healthy person and distributing her organs. Is it ok to cut her up? 13

Trolley A dilemma; also known as *Switch*. A runaway trolley is about to run over and kill five people. You can hit a switch that will divert the trolley onto a different set of tracks where it will kill only one. Is it okay to hit the switch? 13

trolley cases Scenarios designed to elicit puzzling or informative patterns of judgement about how someone should act. Examples include Trolley, Transplant, and Drop. Their use was pioneered by Foot (1967) and Thomson (1976), who aimed to use them to understand ethical considerations around abortion and euthanasia. 5

References

Bago, B. & De Neys, W. (2019). The Intuitive Greater Good: Testing the Corrective Dual Process Model of Moral Cognition. *Journal of Experimental Psychology: General*, *148*(10), 1782–1801.

Bicchieri, C. (2005). *The Grammar of Society: The Nature and Dynamics of Social Norms*. Cambridge University Press.

Boyd, R. & Richerson, P. J. (2005). *The origin and evolution of cultures*. Evolution and cognition. Oxford ; New York: Oxford University Press. [Online; accessed 2025-05-28].

Brennan, G., Eriksson, L., Goodin, R. E., & Southwood, N. (2013). *Explaining Norms*. Oxford University Press.

Butterfill, S. A. & Sinigaglia, C. (2022). Towards a Mechanistically Neutral Account of Acting Jointly: The Notion of a Collective Goal. *Mind*, *132*(525), 1–29.

Chapman, H. A. (2018). *A component process model of disgust, anger, and moral judgment*, volume 70, (pp. 70—80). New York: Guilford Publications.

Chapman, H. A. & Anderson, A. K. (2013). Things rank and gross in nature: A review and synthesis of moral disgust. *Psychological Bulletin*, *139*(2), 300–327.

Chapman, H. A., Kim, D. A., Susskind, J. M., & Anderson, A. K. (2009). In Bad Taste: Evidence for the Oral Origins of Moral Disgust. *Science*, *323*(5918), 1222–1226.

Conway, P. & Gawronski, B. (2013). Deontological and utilitarian inclinations in moral decision making: A process dissociation approach. *Journal of personality and social psychology*, *104*(2), 216–235.

Eskine, K. J., Kacinik, N. A., & Prinz, J. J. (2011). A Bad Taste in the Mouth: Gustatory Disgust Influences Moral Judgment. *Psychological Science*, *22*(3), 295–299.

Foot, P. (1967). The problem of abortion and the doctrine of the double effect. *Oxford Review*, *5*, 5–15.

Gawronski, B., Armstrong, J., Conway, P., Friesdorf, R., & Hütter, M. (2017). Consequences, norms, and generalized inaction in moral dilemmas: The CNI model of moral decision-making. *Journal of personality and social psychology*, *113*(3), 343–376.

Gawronski, B., Conway, P., Armstrong, J., Friesdorf, R., & Hütter, M. (2018). Effects of incidental emotions on moral dilemma judgments: An analysis using the CNI model. *Emotion*, *18*(7), 989–1008.

Giner-Sorolla, R. & Chapman, H. A. (2017). Beyond Purity: Moral Disgust Toward Bad Character. *Psychological Science*, *28*(1), 80–91. [Online; accessed 2023-02-05].

Giner-Sorolla, R., Kupfer, T., & Sabo, J. (2018). What Makes Moral Disgust Special? An Integrative Functional Review. In J. M. Olson (Ed.), *Advances in Experimental Social Psychology*, volume 57 (pp. 223–289). Academic Press.

Greene, J. D. (2007). The Secret Joke of Kant's Soul. In W. Sinnott-Armstrong (Ed.), *Moral Psychology, Vol. 3* (pp. 35–79). MIT Press.

Greene, J. D. (2014a). Beyond Point-and-Shoot Morality: Why Cognitive (Neuro)Science Matters for Ethics. *Ethics*, *124*(4), 695–726.

Greene, J. D. (2014b). *Moral Tribes: Emotion, Reason and the Gap between Us and Them*. New York: Atlantic Books.

Greene, J. D. (2015). The cognitive neuroscience of moral judgment and decision making. In *The Moral Brain: A Multidisciplinary Perspective* (pp. 197–220). Cambridge, MA, US: MIT Press.

Greene, J. D. (2017). The rat-a-gorical imperative: Moral intuition and the limits of affective learning. *Cognition*, *167*, 66–77.

Haggard, P. & Chambon, V. (2012). Sense of agency. *Current Biology*, *22*(10), R390–R392.

Hart, H. L. A. (1994). *The concept of law* (2nd ed ed.). Clarendon law series. Oxford: Clarendon Press.

Heyes, C. M. (2024). Rethinking Norm Psychology. *Perspectives on Psychological Science*, *19*(1), 12–38.

Joyce, R. (2006). *The evolution of morality*. Life and mind. Cambridge, MA ; London: MIT Press. [Online; accessed 2024-10-08].

Koriat, A. (2000). The Feeling of Knowing: Some Metatheoretical Implications for Consciousness and Control. *Consciousness and Cognition*, *9*(2), 149–171.

Koriat, A. (2007). Metacognition and consciousness. In P. D. Zelazo, M. Moscovitch, & E. Thompson (Eds.), *The Cambridge Handbook of Consciousness* (pp. 289–325). New York, NY, US: Cambridge University Press.

Kurzban, R., DeScioli, P., & Fein, D. (2012). Hamilton vs. Kant: Pitting adaptations for altruism against adaptations for moral judgment. *Evolution and Human Behavior*, *33*(4), 323–333.

Lai, C. K., Haidt, J., & Nosek, B. A. (2014). Moral elevation reduces prejudice against gay men. *Cognition and Emotion*, *28*(5), 781–794.

Landy, J. F. & Goodwin, G. P. (2015). Does incidental disgust amplify moral judgment? A meta-analytic review of experimental evidence. *Perspectives on Psychological Science*, *10*(4), 518–536.

MacDonald, G. & Leary, M. R. (2005). Why Does Social Exclusion Hurt? The Relationship Between Social and Physical Pain. *Psychological Bulletin*, *131*(2), 202–223.

Nagel, J. & Waldmann, M. R. (2013). Deconfounding distance effects in judgments of moral obligation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(1), 237.

Ness, J., Mooney, K., & Lach, L. (2010). *Ants as Mutualists*, (pp. 97–114). Oxford, UK: Oxford University Press.

Piazza, J., Landy, J. F., Chakroff, A., Young, L., & Wasserman, E. (2018). What disgust does and does not do for moral cognition. In N. Strohminger & V. Kumar (Eds.), *The Moral Psychology of Disgust* (pp. 53–81). Rowman & Littlefield International.

Richerson, P. J. & Boyd, R. (2005). *Not by genes alone: how culture transformed human evolution.* Chicago: The University of Chicago Press.

Rosas, A. & Aguilar-Pardo, D. (2020). Extreme time-pressure reveals utilitarian intuitions in sacrificial dilemmas. *Thinking & Reasoning*, *26*(4), 534–551.

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

Scott, R. B. & Dienes, Z. (2008). The conscious, the unconscious, and familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(5), 1264–1288.

Suter, R. S. & Hertwig, R. (2011). Time and moral judgment. *Cognition*, *119*(3), 454–458.

Terrizzi, J. A., Shook, N. J., & Ventis, W. L. (2010). Disgust: A predictor of social conservatism and prejudicial attitudes toward homosexuals. *Personality and Individual Differences*, *49*(6), 587–592. [Online; accessed 2025-06-03].

Thomson, J. J. (1976). Killing, Letting Die, and The Trolley Problem. *The Monist*, *59*(2), 204–217.

Tracy, J. L., Steckler, C. M., & Heltzel, G. (2019). The physiological basis of psychological disgust and moral judgments. *Journal of Personality and Social Psychology*, *116*(1), 15–32.

Trémolière, B. & Bonnefon, J.-F. (2014). Efficient Kill–Save Ratios Ease Up the Cognitive Demands on Counterintuitive Moral Utilitarianism. *Personality and Social Psychology Bulletin*, *124*(3), 379–384.

Vanaman, M. E. & Chapman, H. A. (2020). Disgust and disgust-driven moral concerns predict support for restrictions on transgender bathroom access. *Politics and the Life Sciences*, *39*(2), 200–214.

Vlerick, M. (2017). *Better than our nature? Evolution and moral realism, justification, and progress*, (pp. 226—239). Cambridge: Cambridge University Press.

Waldmann, M. R., Nagel, J., & Wiegmann, A. (2012). Moral Judgment. In K. J. Holyoak & R. G. Morrison (Eds.), *The Oxford Handbook of Thinking and Reasoning* (pp. 274–299). Oxford: Oxford University Press.

Wan, L., Dienes, Z., & Fu, X. (2008). Intentional control based on familiarity in artificial grammar learning. *Consciousness and Cognition*, *17*(4), 1209–1218.

Westra, E. & Andrews, K. (2022). A pluralistic framework for the psychology of norms. *Biology & Philosophy*, *37*(5), 40.

Whittlesea, B. W. A. & Williams, L. D. (1998). Why do strangers feel familiar, but friends don't? a discrepancy-attribution account of feelings of familiarity. *Acta Psychologica*, *98*(2-3), 141–165.

Wiegmann, A., Horvath, J., & Meyer, K. (2020). Intuitive expertise and irrelevant options. *Oxford studies in experimental philosophy*, *3*, 275–310.

Wright, R. (1994). *The moral animal: evolutionary psychology and everyday life*. New York: Pantheon Books.