



LUND UNIVERSITY

A Naturalistic Theory of (In)justice

How Neurophysiology and Metabolic Energy Ground the Perception of Injustice

Mirzaeighazi, Shervin; Tjøstheim, Trond A.

Published in:

Journal for the Theory of Social Behaviour

DOI:

[10.1111/jtsb.70027](https://doi.org/10.1111/jtsb.70027)

2026

Document Version:

Publisher's PDF, also known as Version of record

[Link to publication](#)

Citation for published version (APA):

Mirzaeighazi, S., & Tjøstheim, T. A. (2026). A Naturalistic Theory of (In)justice: How Neurophysiology and Metabolic Energy Ground the Perception of Injustice. *Journal for the Theory of Social Behaviour*, 56(1), Article e70027. <https://doi.org/10.1111/jtsb.70027>

Total number of authors:

2

Creative Commons License:

CC BY

General rights

Unless other specific re-use rights are stated the following general rights apply:

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: <https://creativecommons.org/licenses/>

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

ORIGINAL ARTICLE OPEN ACCESS

A Naturalistic Theory of (In)justice: How Neurophysiology and Metabolic Energy Ground the Perception of Injustice

Shervin MirzaeiGhazi  | Trond A. Tjøstheim 

Department of Philosophy and Cognitive Science, Lund University, Lund, Sweden

Correspondence: Shervin MirzaeiGhazi (shervin.mirzaeighazi@fil.lu.se)

Received: 16 June 2025 | **Revised:** 10 November 2025 | **Accepted:** 29 December 2025

Keywords: energy | evolution | injustice | neurophysiology | perception | theory of mind

ABSTRACT

Across different domains, justice is considered either from a perspective concerning mind-independent features of a situation or from a perspective related to mind-dependent motives, traits or emotions. Although these approaches have generated valuable insights, they remain largely disconnected from each other. What is missing is an integrative framework that can explain how the objective features of situations are translated into subjective experiences and, ultimately, into defensible moral judgements. In this paper, we propose a physiological account of justice perception that explains how the ontologically objective features of a situation are translated into epistemically objective moral judgements. We argue here that the brain's predictive processing and interoceptive systems register unwarranted energy costs imposed by others as a salient, ontologically subjective state of negative affect through prediction errors elicited in specifically social contexts. This affective signal, in turn, provides the motivational and phenomenological basis for the cognitive judgement—mediated by Theory of Mind—that constitutes a specific moral emotion, such as indignation or outrage. Our naturalistic framework thus illuminates the embodied foundations of justice without reducing moral judgement to a mere physiological response. Although we are proposing a neurophysiological mechanism for justice perception, the content of the cognitive judgement (what counts as 'unjustified' or a norm of 'fairness') is also shaped by cultural, sociological and anthropological context. Thus, rather than replacing existing philosophical, legal or psychological theories, this naturalistic framework complements them by revealing the embodied cognitive processes that underlie our justice judgements across diverse contexts and cultures.

1 | Introduction

Although justice norms differ across societies, all known cultures develop moral systems and frameworks for justice, indicating that the underlying motivation and concern are universally present (Haidt 2012). As Gollwitzer and van Prooijen (2016, 66–67) claim, the general concern with justice and morality among different people and cultures can be considered as empirical evidence of a universal concern for justice. Many researchers in psychological justice agree that assuming all humans have an inborn concern for justice helps both in understanding human behaviour (theory) and in solving practical problems (Ellard et al. 2016).

Evolutionary approaches have sought to understand how these inborn determinants of justice could contribute to the overall fitness of human beings and how creating more or less just societies could increase the survivability of individuals and human groups. These approaches provide crucial context for understanding why humans across cultures share certain basic intuitions about justice despite significant variation in specific norms and institutions. According to them, our sense of fairness may have evolved as an adaptation to facilitate cooperation in social groups (De Waal 1996; Van Vugt et al. 2008; Boehm 2012). Brosnan and de Waal's (2003) research on inequity aversion in nonhuman primates indicates that the foundations of justice perception may predate human societies. Capuchin monkeys

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2026 The Author(s). *Journal for the Theory of Social Behaviour* published by John Wiley & Sons Ltd.

and chimpanzees have been observed rejecting unequal rewards, suggesting that sensitivity to fairness has deep evolutionary roots.

This evolutionary perspective is supported by a wide range of scientific findings highlighting the importance of justice in everyday social interactions. Justice concerns have been shown to be vital not only in ancestral environments but also in contemporary settings such as organisations, legal systems, classrooms, sports teams, friendships and intimate relationships. Studies in these domains consistently reveal that fairness contributes to social cohesion, trust and satisfaction (De Cremer and Tyler 2005; Tyler and Blader 2000; Tyler and Lind 1992; Van den Bos and Lind 2002).

Evolutionary game theory models have demonstrated how punishment of free riders can stabilise cooperation (Fehr and Gächter 2002), suggesting that sensitivity to justice violations may have conferred fitness advantages in ancestral environments. This finding supports the idea that by promoting cooperation and norm enforcement, punishment serves a critical social function. Relatedly, a meta-analysis by Balliet et al. (2011) demonstrates that punishment systems generally have a positive effect on cooperation and reciprocity across various social settings. The ability to detect and respond to unfair treatment would have been particularly valuable in small-scale societies, where repeated interactions and reputation management were crucial for survival.

These accounts, although insightful, are incomplete. Any complete theory of justice must deal with a fundamental puzzle: How do human beings move from observing a set of facts in the world to making a robust, defensible moral claim about those facts? This transition is central to all justice-related debates, yet its underlying mechanism is not fully fleshed out. To clarify this problem, it is helpful to use a philosophical distinction between claims about existence (ontology) and claims about knowledge (epistemology).¹ An event or state of affairs can be *ontologically objective*, meaning that its existence is independent of any mind or observer, such as mountains and molecules. On the other hand, pains, itches and the specific subjective experience of a colour (qualia) are *ontologically subjective*, meaning that their mode of existence is mind-dependent. These entities exist only insofar as they are experienced.

The same distinction exists concerning the status of our *judgements or claims* about the world. The truth or falsity of a claim can be settled as a matter of fact, independent of any individual's feelings, attitudes or point of view. The statement, 'Paris is the capital of France', is *epistemically objective*. Its truth can be verified or falsified through historical evidence. The truth or falsity of a claim, however, can depend on the attitudes, feelings or preferences of the person making it. The statement, 'Paris is more beautiful than London', is *epistemically subjective*. It is a matter of taste or opinion, not a verifiable fact. Crucially, these two distinctions are independent. As Searle (1995) notes, one can make epistemically objective statements about ontologically subjective entities. For instance, the statement 'I am currently in pain' is a claim about an ontologically subjective state (the pain itself), but the claim can be epistemically objective (it is a verifiable fact that I am, or am not, in that state).

A moral judgement, such as 'That distribution is unjust', aspires to be epistemically objective; its truth or falsity is not merely a matter of personal taste but can be settled through reason, evidence and shared principles. The challenge is thus to explain the process that connects the ontologically objective features of a situation—for instance, a measurable and verifiable disparity in resource allocation—to an epistemically objective moral claim about its fairness. This challenge is in line with a broader shift in cognitive science towards understanding the mind as embodied, embedded, extended and enactive (often termed 4E cognition).² In particular, according to Varela, moral understanding is not just the application of abstract rules, but a form of skillful engagement with the world (Gallagher 2024). This perspective, which is usually called enactivist ethics, emphasises 'ethical know-how'—a situated, affectively engaged, and immediate coping that characterises moral expertise (Hutto 2025).

However, although enactivist ethics provides a powerful philosophical framework for understanding the embodied nature of moral experience, like many other philosophical and psychological accounts (of morality and justice), it has often remained at a high level of description. What has been largely missing is a neurobiological mechanism that explains *how* a situation becomes ethically salient to an organism. The feeling of being wronged, the physiological stress and the negative affect associated with perceived unfairness are all ontologically subjective states. We claim that this subjective experience is the critical link between worldly facts and moral judgement. We argue below that this experience is grounded in a physiological aversion to unwarranted energy expenditure, a process mediated by the brain's interoceptive and predictive processing systems. In this way, our neurophysiological theory complements normative theories of justice by providing an explanation of the causal mechanism that makes them psychologically and motivationally effective.

To start with, social and affective neuroscience have indicated that motivation for seeking justice has both an intuitive, heuristic and fast component, as well as a slower and more deliberative one (Decety and Yoder 2017). At the same time, people are differently motivated to seek justice, and the motivation to do this on the part of oneself versus on the part of others appears to be independent. These motivations may even be in opposition. In other words, there are both idiosyncratic differences in people's propensity to engage in justice-seeking behaviour, as well as their propensity to do this either on their own behalf, or someone else's. Although neural correlates have been established for where in the brain evaluations of fairness, justice and moral judgement appear to be mediated (Ruff and Fehr 2014, for a review), it is still unclear what drives activity in the relevant areas.

To avoid biological reductionism, we propose a two-part model of moral emotion that distinguishes its affective and cognitive components. The neurophysiological mechanism of detecting unwarranted energy costs accounts for the *affective component* of our moral responses—the raw, visceral and ontologically subjective feeling of wrongness that gives moral emotions their motivational force. However, this raw affect is not, in itself, a moral emotion like outrage or indignation. It is interpreted and transformed into a specific moral emotion through a *cognitive*

judgement about the situation, particularly a judgement about the intentions of the agents involved. This cognitive layer, mediated by processes such as Theory of Mind, explains how the same underlying physiological signal can be experienced as misfortune in one context and as injustice in another.

The main contribution of this paper is thus to propose an embodied account of justice perception that explains how the ontologically objective features of a situation (i.e., the uncompensated energy costs imposed on someone) are translated into epistemically objective moral judgements (i.e., the other is acting unjustly). We argue here that the brain's predictive processing and interoceptive systems register unwarranted energy costs imposed by others as a salient, ontologically subjective state of negative affect. This affective signal in turn provides the motivational and phenomenological basis for the cognitive judgement—mediated by Theory of Mind—that constitutes a specific moral emotion, such as indignation or outrage. Our naturalistic framework illuminates the embodied foundations of justice without reducing moral judgement to neurophysiological responses alone.

The remainder of this paper develops this account in six sections. Section 2 establishes the fundamental premise that humans are averse to energy expenditure, particularly when those expenditures are unexpected. Section 3 explores how interoception—the brain's monitoring of bodily states—mediates this aversion by detecting energy expenditure and generating affective responses. Section 4 distinguishes between different types of energy expenditure, explaining why some forms are tolerated while others trigger negative responses. Section 5 introduces the central roles of agency detection and cognitive judgement in transforming a negative affective state into a moral emotion. Finally, Section 6 discusses how our account complements and illuminates traditional intuitions about justice.

2 | The Physiological Basis of Our Aversion to Energy Expenditure

Humans, like all other living organisms, have evolved to conserve energy. This aversion to unnecessary energy expenditure is rooted in our physiology. The body carefully regulates energy through metabolic processes and ensures that we do not waste resources on activities that do not contribute to our survival or well-being. All cells in the human body are dependent on sufficient energy to maintain their structural integrity and do their job, whether it be to protect the body as part of the immune system, to inform the central nervous system about the external world or to process that information in the brain to make decisions. Hence, our behaviour, from cooperating socially with others to calculating how much money we owe them, is profoundly affected by energy.³

2.1 | Homeostasis and Energy Conservation

The process of predictively maintaining the structure and functions of cells is referred to as *allostasis*; the idealised state of the cell being at its optimum function is called *homeostasis* (Cannon 1932). In practice, cells are continuously affected by

dynamic changes of the internal environment and are thus seldom in *homeostasis* for long. However, as long as the changes are not too large and the supply of energy is sufficient, a collection of feedback processes will cooperate to compensate for the perturbations. This is easier to understand on the levels of the organism as a whole: If you go into a cold shower, the body will compensate for the drop in temperature by shivering, whereas on a sunny day, the increase in temperature will be compensated by sweating.

2.2 | Neurological and Hormonal Regulation

The brain, particularly regions like the hypothalamus, plays a key role in monitoring energy reserves and signalling when conservation is necessary (Kandel et al. 2021, ch 41). These regions integrate with the nervous system of the gut, and hormones such as leptin and ghrelin regulate hunger, satiety and energy balance, influencing motivation and effort (Menzies et al. 2013; MacCormack and Muscatell 2019). More specifically, ghrelin is involved with hunger, whereas leptin contributes to satiation (MacCormack and Muscatell 2019); both ghrelin and leptin are active even in social contexts that are related to *allostasis*. In other words, both hormones contribute to predicting energy-related outcomes that are affected by relations to other people, such as social support and social status. Both loss of social support and reduction of social status can, in this way, be viewed as events that predict significant potential energy deficits and tend to be strongly aversive.

It must be mentioned that humans are not universally resistant to all forms of energy expenditure. There are instances when we experience an intrinsic drive to engage in certain energy-costly activities, even when these actions are not immediately motivated by a clear, instrumental goal, such as exercising for the purpose of improving health. This observation, however, does not contradict the broader claim being made here. Our argument focuses on a fundamental physiological mechanism that predisposes individuals to avoid unnecessary energy consumption. Yet, the existence of such a mechanism does not preclude the possibility that other physiological processes might, under specific conditions, encourage energy expenditure (a classic example is play behaviour, see, e.g., Takhvar 1988). A plausible hypothesis is that the inclination to expend energy could be linked to periods of energy abundance. In such conditions, an increased tendency towards movement and activity might confer an evolutionary advantage, either at the individual or group level. For instance, in an environment with a large amount of resources and minimal predation, animals that are more inclined to move may benefit in several ways: They might explore and colonise new areas more effectively, secure additional resources or maintain better physical health. When considering the broader trajectory of evolutionary history, such favourable conditions appear to have been the exception rather than the norm. If this perspective is accurate, it follows that the default biological tendency would be a stronger aversion to energy expenditure than an innate drive towards it.

In this section, we considered the general aversion of the body towards energy expenditure. To be able to avoid energy shortcomings though, the body needs to have a good picture of the

present state of affairs. In the next section, we consider how the body can sense and perceive its own state.

3 | Interoception: Sensing the Body's Energy State

Interoceptive pathways provide the brain with constant updates on the body's internal state, including energy levels, stress indicators (heart rate, breathing, etc.) and pain. This interoceptive pathway, by translating metabolic states into neural signals, provides the raw material for the *affective component* of our emotional responses to injustice—the felt, negatively valenced bodily state or *ontologically subjective experience*—that gives the perception of being wronged its visceral power and motivational urgency. The point of this in the context of emotions in general, and moral emotions in particular, is that it is likely that even abstract moral notions like justice are grounded in processes going on in the body, and more specifically, processes having to do with metabolism and energy availability. The reason for this is that evolution often modifies already existing processes rather than coming up with entirely new systems from scratch (Jacob 1977). Metabolism is a fundamental and ancient biological process (e.g., Craig 2002) that is directly linked to energy and is a plausible candidate process for being co-opted by evolution to be involved in more complex processes that involve energy as well.

Social competition, cooperation and social support all involve some form of energy accounting (Beckes and Coan 2011), and interoceptive signals regarding predicted energy costs, like those associated with threats or savings involved in cooperation and support, are valuable information sources when making decisions regarding, for example, fight or flight. However, this innate grounding does not preclude the need for learning and adaptation. Hence, processes related to moral notions, for example, fairness and unfairness, can also be learnt through interaction with others and immersion in a culture, complementing the intuitive signals. In other words, there is an intuitive aspect of feelings of injustice, the immediate recognition of a norm violation when somebody directly and unexpectedly harms you physically (Hamlin 2013). Things like queueing rules in supermarkets may never even be spoken out loud but will be learnt through observing and feeling others' protests when someone violates the norm, for example, by trying to skip the line.

Interoception serves as a bridge between physiological processes in the body and cognitive processes by influencing emotions and decision-making. In physical bodies, interoceptive information from the body's internal sensors flows through the interoceptive pathway (Craig 2002), which includes the spinal cord, the brainstem, the thalamus and cortical areas like the anterior insula (AI) and the anterior cingulate (ACC). As we will see below, the latter two areas are particularly interesting for perceptions of injustice and decisions to punish norm violators. Inequity can have a direct and affective effect on a person, eliciting emotions like anger and resentment (Batson et al. 2007). For example, if someone causes you physical pain by slapping you, you might react immediately by becoming angry and wanting to hit back; this kind of defensive reaction is fast and does not involve a lot of consequence thinking. Higher-level

concepts like inequity and injustice, on the other hand, require the integration of many interoceptive signals and may draw on cognitive processes like memory (Goldinger et al. 2003) and Theory of Mind (ToM; Young et al. 2007)—the ability to take the perspective of another agent and make judgements based on the perceived value system of that other.

On a physiological level, the infliction of harm on a person activates that person's stress response. Fundamentally, the stress response is closely connected to energy allocation since a stressor is typically something that increases the body's energy requirements for maintaining its integrity (Peters et al. 2017; Sterling 2004). This happens through a complex coordination of bodily processes, including a rise in cortisol and adrenaline levels, which in turn speed up the heart and breathing and draws on the body's energy reserves. Depending on the circumstance, the stress response also primes the behavioural system to either engage with or escape the stressor (both energy-costly processes)—commonly referred to as the fight-or-flight response. This behavioural readiness can manifest not only in flight or direct aggression, but also in socially mediated forms of engagement, such as punitive or protective actions. Such actions, though still energetically costly, extend the physiological stress response into the social domain, where perceived threats to social integrity (e.g., unfairness or harm to others) elicit responses aimed at restoring balance. For example, human toddlers exhibit a readiness to punish anti-social agents even as uninvolved third parties (Hamlin et al. 2011), and they show expectations that others will intervene to protect victims (Geraci 2021), suggesting an early emerging link between physiological arousal, perceived injustice and social enforcement behaviours.

Although organisms possess a fundamental drive to conserve energy, our subjective experience of expending energy is not uniformly negative. In the next section, we expand on the factors that differentiate negatively perceived energy costs from positively or neutrally experienced energy investments.

4 | The Dual Nature of Energy Expenditure: Positive Versus Negative Experiences

4.1 | The Aversive Nature of Uncompensated Energy Costs

To briefly recap, allostasis ensures that the body predicts future energy requirements relative to the current metabolic state. This means that we can become more sensitive to a cost versus payback calculus if we are hungry, but are more motivated to spend energy if we feel well fed or that we will be sufficiently compensated for any energy outlays.

The core idea in this section is that negative experiences tend to arise from energy expenditure being perceived as uncompensated, unexpected or forced. An example of a quintessential negative experience is pain. Pain can be connected to metabolic energy first via the pure metabolic cost of neuronal excitation in the pain pathways (Waxman et al. 1999) and second via engagement of the body's stress response. Pain typically signals a threat to the body's integrity, as pain sensors (called nociceptors)

are activated when tissues are harmed in various ways (Dubin and Patapoutian 2010), and thus become in need of repair or defence. In this sense, the energy expense is inherently negative and draining (it opposes allostasis and growth processes).

Looking at the information signalling involved in pain and stress from the perspective of predictive processing (e.g., Clark 2016; Hohwy 2013), large prediction errors, such as those involved when the body is hurt or attacked, might themselves be inherently painful or aversive (Kiverstein et al. 2022). The nature of how prediction errors are coded in the nervous system implies that neurons become excited when prediction errors occur, and neural excitation costs energy. This cost can thus be connected to unexpected energy demands, that is, situations where the brain's predictive system fails to anticipate costs, which leads to an aversive mismatch signal. An unexpected injury would cause such a signal, but so could potentially a social situation where, for example, an unexpected violation of fairness would demand a costly response in terms of protest or attack. Pain, stress and fatigue are all *felt* through interoceptive pathways, indicating that they are a substantial part of the raw negative valence (Craig 2002).

4.2 | Energy Expenditure as a Tolerated Necessity or Valued Investment

Despite our default aversion to energy costs, we often actively seek them out through sports and play. Why might this be so? Reconnecting to Section 2 above, there is sometimes an urge to spend energy to return to homeostasis. Similarly, motivation can be strong to reduce classic motivational drives like hunger and thirst, even if foraging for food and water can imply substantial effort. Foraging costs like this can be framed as necessary expenditures driven by interoceptive deficit signals (i.e., hunger and thirst).

When foraging is successful, there is also a reward waiting, namely the pleasure of consummation when eating and drinking. From an allostatic perspective, the positive experience of reducing hunger or thirst can be linked to the intrinsic reward of restoring homeostasis. This makes the effort both tolerable as a necessity to uphold life, but also a kind of valued investment when foraging is done on a full belly, so to speak. In the latter case, it helps ensure that future needs will be met.

To clarify further, the positive experience of reducing fundamental drives like hunger can be divided into the reduction of *wanting* food and the consummatory *liking* of eating (Kringelbach 2010). 'Wanting' and 'Liking' are technical terms that involve complementary brain systems, namely the dopamine system and the opioid system. The dopamine system can be thought of as being responsible for seeking out and approaching food, whereas the opioid system is responsible for initiating, maintaining and ending the process of eating (Le Merrier et al. 2009). In the current context, the relevance lies in the capacity of both systems to *modulate* the energy involved in neuronal activation by changing neuronal activity thresholds (dopamine; Beaulieu and Gainetdinov 2011) and being associated with low-arousal positive states and post-consummatory 'calm' modes (opioids: e.g., Gosnell and Levine 2009; Kringelbach 2010).

4.3 | Interoception Modulated by Expectation and Meaning

Although energy expenditure can be similar while being experienced in completely opposite ways—negative when unexpected or forced, positive if intrinsically motivated—context and prediction clearly modulate which one it will be. In other words, how does the brain differentiate the cost of work during forced labour versus voluntary exercise? Or the stress of a surprise attack versus the anxious excitement before a competition?

We propose that top-down cognitive appraisal, integrating expectations, perceived control and perceived fairness or meaningfulness of the task, modulates interoceptive signals associated with energy expenditure. That is, the overall *meaning* of the interoceptive signals affects whether the signals are subjectively experienced as positive or negative. More concretely, the brain produces different predictions depending on its appraisal, radically changing what *produces* costly prediction errors, and whether prediction errors are tolerated as necessary, valued as investment or welcomed as contributing to a greater cause, rather than abhorred as unfair and unjust. This modulatory role of cognition is critical for understanding moral emotions, where a specific *cognitive judgement* about an agent's intentions can transform the very same affective signal from an experience of mere loss or misfortune into one of injustice, a process we will elaborate on in the following section.

To summarise, we can experience energy costs as neutral or positive if they are appraised as necessary, as an investment or if we have excess energy that needs to be spent. When we do experience them as negative, it is often because mitigating factors like expected compensation, control or inherent meaning are absent. In the next section, we will go further into two specific forms this negative experience can take in contexts that are relevant to fairness and justice.

5 | The Two Forms of Negative Energy Expenditure and Their Emotional Consequences: Nonagentic Loss Versus Perceived Intentional Injustice

In the preceding sections, we established that humans exhibit a general aversion to energy expenditure, particularly when it is imposed unexpectedly or without compensation, resulting in a negatively valenced, ontologically subjective experience. This section develops a necessary distinction in how these negative experiences are structured into specific emotions. To achieve this, we follow Solomon's (1988) theory of emotion, according to which an emotion is not merely a passive feeling but a form of active, evaluative engagement with the world. An emotion is constituted by both an *affective component* (the feeling or physiological sensation) and a *cognitive component* (an evaluative judgement about the situation). The physiological processes described thus far—the interoceptive detection of an unexpected energy cost—again provide the ingredients for the affective component. This section explains the cognitive component: the judgement that transforms the raw affect into either a nonmoral negative emotion, like frustration, or a specific moral emotion, like outrage. The key variable in this transformation is the

perceived cause of the energy cost, specifically whether it is attributed to an intentional agent.

5.1 | Negative Expenditure via Nonagentic Causes—Loss, Futility and Misfortune

For the purpose of the current work, we define ‘non-agentic causes’ as including misfortune, disappointment due to impersonal circumstances, systemic barriers perceived as difficult but not personally malicious or failing to achieve a goal despite personal effort (i.e., futility). These are causes that can be attributed to bad luck, environmental difficulties, one’s own limitations and so on, rather than active opposition and malice from another agent. In these instances, the negative affect stemming from the unexpected energy cost is present, but the specific cognitive judgement of intentional, compensable harm is absent.

The prediction errors generated by these impersonal forces thus relate to how predictable the environment is, how probable it is in general to attain the specific goal and how probable it is for *you* to attain the goal (i.e., your self-efficacy). Hence, in these cases, there is no betrayal of social trust or of any fairness norms. Consequently, the typical emotional response to any setback tends to be nonmoral negative emotions, like frustration, sadness and disappointment. For example, a farmer whose crops are destroyed by a hailstorm experiences a significant and unexpected energy loss. The interoceptive signals of this loss generate a negative affective state. However, because the cause is nonagential, this affect is structured by a cognitive judgement about misfortune, not injustice. The resulting emotion is sadness or despair, not outrage.

The behavioural consequences of these nonmoral negative emotions may include efforts to learn from mistakes, change strategy, alter the goal or adapt to the circumstances. If the odds of success are perceived as too low, resignation may also be an outcome. But in general, there tends to be no moral condemnation or demand for punishment directed at any specific entity. The experience is unfortunate, but it is not an injustice. This highlights that the physiological response to an energy cost is a necessary but not sufficient condition for a *moral* emotion.

5.2 | Negative Expenditure via Perceived Intentional Imposition—Perceived Injustice

In contrast to the previous section, we define ‘intentional imposition’ to include the perception of an *agent* as deliberately imposing a cost (or preventing a deserved gain) without justification or adequate compensation. This includes theft, deception, coercion, exploitation, betrayal or the unfair harm caused by another’s action or deliberate inaction. In these circumstances, the interoceptive signal of an unexpected energy cost is the ‘stress’ (the affective component). The cognitive judgement, mediated by Theory of Mind (ToM), that this cost was imposed by a malicious agent for unjust reasons, is the ‘outrage’ (the cognitive component). Together, they form the complete moral emotion.

The role of ToM is to facilitate the *cognitive appraisal* or *judgement* that constitutes the emotion. This judgement is not

something that follows the feeling. It is the cognitive act of judging that one has been intentionally and unfairly harmed that *structures the negative feeling as* the specific emotion of indignation or outrage. The prediction errors in these cases are not about the physical environment but about violations of expected social norms, fairness, trust and the predicted behaviour of the other agent—that is, they are specifically *social* prediction errors. The emotional responses are qualitatively different from nonagentic cases and include distinctively *moral* emotions such as resentment, indignation, outrage or contempt, which are elicited in response to perceived unfairness, power abuse or betrayal (Haidt 2003). Although nonmoral emotions are responses to harm *qua* harm (like sadness), moral emotions are responses to compensable harms—an earthquake cannot compensate for the damage it inflicts, but a human agent can.

Behaviourally, these negative moral emotions motivate demands for justice, accountability, punishment, retribution or compensation.⁴ There is a motivation to confront or avoid the perpetrator, to protest, make demands or withdraw social support. This behaviour together reinforces the moral characteristics of the event and contributes to the development and enforcement of social norms and judicial systems such that future transgressions become less likely.

This distinction between the raw affective component and the structuring cognitive judgement maps directly onto a key debate within the enactivist theory concerning the relationship between immediate, ‘contentless’ bodily coping and ‘contentful,’ deliberative reasoning (Hutto 2025). Our model provides a neurophysiological account of this integration. The initial interoceptive signal—the detection of an unexpected energy cost—represents the immediate, pre-reflective, *contentless* bodily coping that Varela identified. It is the raw, visceral feeling that something is wrong. However, this signal alone is insufficient for a full moral judgement. The second stage of our model—the ToM-mediated cognitive appraisal of intentionality—provides the *contentful* layer. This cognitive act structures the raw affect into a specific, justifiable moral emotion (e.g., indignation) and forms the basis of an epistemically objective claim (‘That was unjust’).

5.3 | The Decisive Role of Attributed Intentionality—Theory of Mind

The distinguishing factor between the two types of negative experience described above is the perceived presence of *intentionality*, not the *magnitude* of the energy cost. Even a large setback in effort versus gain, such as when a newly built house is damaged due to an intense storm, will sooner elicit sadness over the loss rather than indignation. Conversely, a slight, like withholding a small amount of payment or refusing to pay up for lunch, can produce intense resentment if the victim perceives significant ill intent (rather than justifiable ignorance, say) as the cause. The energy cost of the latter may be minor, but the cognitive judgement of intentional harm transforms the experience into a moral violation.

In this way, ToM acts as a cognitive switch, toggling the interpretation of an energy cost from a ‘loss’ or ‘misfortune’ to an

‘injustice’ or ‘violation’. Although both interpretations involve negative valence, only with the attribution of intentionality does the negativity become a *moral* issue. Furthermore, once this switch has been made, it opens the door for additional moralisation regarding blame (who was the culprit), responsibility (who must make amends) and social judgement (e.g., will group membership and social support be withdrawn?).

To summarise, energy costs may be evaluated negatively independent of their perceived causes, but they are associated with qualitatively different emotions depending on whether those causes involve intentionality and agency. In the latter case, ToM may act as a switch, changing the interpretation and subsequent behaviour to involve *moralising* and activating concepts such as fairness, blame and responsibility. The heightened negative valence and specific moralised emotions elicited by perceived intentional harm point towards a deeper functional significance.

6 | Discussion

6.1 | Taking Stock

In this work, we have presented an account of how humans perceive injustice through a mechanism that transforms basic biological energy conservation instincts into complex moral judgements. The theory proposes a two-part (affective + cognitive) model that connects different levels of analysis. The central insight of our theory is that violations of fairness feel deeply wrong because they register physiologically as threats to our energy resources. This provides the *affective force* and *motivational urgency* for our moral responses. This process begins with the interoceptive detection of an unexpected or uncompensated energy cost, which generates a negatively valenced, ontologically subjective state.

However, our theory is nonreductionist. This raw affective signal does not, on its own, constitute a moral experience. A higher-order judgement, mediated by Theory of Mind, provides the cognitive structure of a moral emotion, which appraises the cause of the energy cost. When this cost is attributed to the intentional and unjustified actions of another agent, the raw negative affect is transformed from an experience of simple loss or misfortune into a specific moral emotion like indignation or outrage. This critical cognitive switch transforms what could have been a personal problem into a moral violation, generating justice-seeking behaviours like confrontation, punishment or norm enforcement. This two-part model grounds the experience of justice in the concrete reality of embodied existence while simultaneously accounting for the central role of cognitive appraisal, cultural learning and social context in shaping our moral judgements.

The emotional and punitive responses to injustice described in our physiological account have likely been evolutionarily selected for their benefits in facilitating cooperation. Many scholars highlight that the emotional urge to punish wrongdoers is probably a hardwired instinct that evolved because it effectively supports strategies like deterrence, even if we do not consciously calculate the consequences every time we feel it (Carlsmith et al. 2002; Hoffman and Goldsmith 2004; Gintis et al. 2008).

Natural selection may have programmed us with a retributive instinct because that instinct leads to behaviours that maintain social order. The reflex to defend against an attacker by means of aggression has a deep evolutionary history and appears to go beyond humans and even mammals (Thomas et al. 2015). Many species, from insects (Wenseleers and Ratnieks 2006) to fish (Bshary and Grutter 2005), demonstrate behaviour similar to second-party punishment—the retaliation of an insulted party against a violator. Since exploitation can trigger punishment (Dong et al. 2022), and people show an aversion to being exploited (Carpenter and Dolifka 2017), this points to strong evolutionary pressures to conserve energy and defend against uncompensated energy costs (Ledoux 2021; Cushman 2014). Second-party punishment may thus have emerged as a solution to this problem (Haidt 2007; Hamilton 1964a, 1964b; Trivers 1971).

Third-party punishment (3PP), where an individual who is not directly affected by a violation still engages in punitive action, may have grown from second-party punishment (Bellucci et al. 2020; Buckholtz and Marois 2012; Marlowe 2009; Marlowe et al. 2011). The propensity for 3PP is observed even in toddlers (Hamlin et al. 2011; Geraci 2021) and reflects an early evolved capacity to perceive and act on such imbalances. By spreading the burden and risk of punishment across the group, 3PP contributes to upholding social norms that inhibit costly social transgressions. This behaviour can be traced to individual urges that have likely been selected for their deterrence value (Carlsmith et al. 2002; Gintis et al. 2008; Hoffman and Goldsmith 2004; Haselton and Buss 2009).

Retributive motivations may reflect a ‘hardwired’ solution to the collective action problem of enforcing norms: Natural selection favoured individuals who risked immediate energy expenditures to punish transgressors, thereby securing long-term group benefits (Haselton and Buss 2009). The long-term fitness benefits of punishment—reduced conflict, maintenance of group cohesion and deterrence of future harm—outweigh its immediate physiological and energetic costs. Moreover, because reputation is a crucial resource in primate social systems (Manrique et al. 2021), enforcing norms through punitive or protective action helps defend one’s standing in ways almost as vital as defending one’s own bodily integrity (Jordan et al. 2016).

6.2 | Applications to Philosophical Intuitions About Justice

Having established the physiological mechanism, we can now examine how our account aligns with, and illuminates, widely held philosophical intuitions about justice. This analysis demonstrates that the energy-based framework not only explains the psychological experience of injustice but also provides a naturalistic foundation for normative principles that philosophers have long recognised as central to justice.

6.2.1 | Justice as Conflict Resolution and Entitlement Allocation

A fundamental intuition held by many philosophers is that justice becomes salient when individuals or groups make

competing claims to resources, freedoms or opportunities (Rawls 1971; Nagel 2005). When resources are abundant or goals are aligned, the relevance of justice often diminishes (Hume 1975, 183–184). Our theory captures this by defining injustice as the intentional imposition of unexpected or uncompensated energy costs by one agent onto another. When someone's freedom or access to resources is restricted, the affected person must expend extra energy—emotionally, physically or economically—to cope with the deprivation or restore what was lost. This unequal distribution of energy burdens, imposed without consent, triggers the experience of injustice.

For instance, being denied a job due to discrimination (an ontologically objective event) forces the individual to expend more energy seeking alternative employment or challenging the injustice. This unchosen and uncompensated energy burden is registered via interoception as a negative affective state (an ontologically subjective experience). When this state is cognitively appraised as stemming from an intentional, discriminatory system, it forms the basis for an epistemically objective judgement that the situation is unjust. Conversely, in conditions of abundance or cooperation, where energy costs are shared, there is no involuntary transfer of energy burdens, and justice is less likely to be a primary concern. The invocation of justice thus emerges precisely when relational imbalances in energy costs arise.

6.2.2 | Justice as Equal Treatment in the Absence of Relevant Differences

One common philosophical principle holds that justice requires treating people equally when there are no morally relevant differences between them (Sen 1980; Cohen 1989; Dworkin 2000).⁵

This principle is especially important in distinguishing between inequalities that arise from luck—things beyond a person's control, such as race, gender or inherited wealth—and those that stem from personal choice. The former is generally seen as unjust and deserving of corrective measures, whereas the latter can be accepted as fair. The theory presented here understands these inequalities in terms of discrepancies in the amount of effort required to achieve the same results. When individuals are treated unequally due to morally irrelevant factors, they are forced to expend significantly more energy to reach the same goals as their more advantaged counterparts.

This process can be analysed with the current framework in the following way. Unequal treatment based on factors like race or gender imposes an ontologically objective and uncompensated energy cost on disadvantaged individuals. They must exert more effort to overcome implicit bias or institutional barriers. The ontologically subjective experience of this physiological burden (e.g., stress, fatigue), when cognitively appraised as stemming from an agential cause, broadly construed, forms the basis for an epistemically objective judgement that the treatment is unjust. In contrast, when a discrepancy in energy cost is expected due to a relevant difference in merit or choice, it is not considered unjust. Legal and institutional commitments to equal treatment can thus be seen as practical tools for ensuring that no group is

forced to carry an unfair share of the social and economic energy burden.

6.2.3 | Power, Status and Justice

In complex societies, the institutionalised system of command and control, which is necessary for large-scale cooperation, inevitably creates hierarchies because some people become more powerful or more respected—that is, they have higher status (Fiske et al. 2016). According to Blader and Chen (2012), power and status hierarchies can be distinguished based on how they influence justice behaviour in individuals who possess either of these attributes. They conducted experiments to examine the impact of power and status on people's enactment of distributive and procedural justice. Their studies showed that having power generally has a negative impact on how people practice justice towards others, whereas status tends to have a positive effect on justice behaviour. The model of justice presented in this paper provides a straightforward explanation for this discrepancy.

One key observation is that having power allows individuals to delegate energetically costly tasks to others without having to reciprocally take on those energy burdens. As Blader and Chen observe, power shifts attention to the self and one's own goals, making people more likely to see others as tools to achieve those goals. In other words, people with power become less inclined to focus on the energy others must spend to support their interests. This reduces the likelihood of justice behaviour in practice. In contrast, status depends more on others—you cannot confer status on yourself. To gain status, people must pay attention to the needs, goals and energy levels of others instead of bossing them around. The interaction between power and status suggests that individuals motivated by both, or primarily by status, are more likely to act justly. Conversely, those with power but low status may be more prone to unethical and unjust behaviour (Fast et al. 2012).

6.2.4 | Conservative Versus Reformist Aspects of Justice

There is another important philosophical intuition that distinguishes between two aspects of justice: one that conserves existing rights and norms and one that pushes for reform based on ideal principles (Sidgwick 1874/1907; Raphael 2001). Our framework offers a naturalistic explanation for this tension by framing it in terms of energy costs. Established norms and institutions create social stability in which individuals develop routines and expectations that minimise energy expenditure. Any reform, even when meant to make the system fairer, comes with new energy costs for those who are used to the old ways. These costs might include learning new rules, adapting to different roles or adjusting to shifts in power dynamics. As a result, people often resist reform not necessarily because they oppose justice in principle but because change requires effort, uncertainty and potential loss of previously enjoyed advantages.

However, the same framework also explains why reform is often demanded in the first place. Existing institutions may themselves be perceived as structurally unjust if they allocate energy burdens

unequally. Marginalised groups—such as women, racial minorities or people with disabilities—often have to bear extra physical, emotional or financial burdens just to compete on equal terms. This can mean dealing with discrimination, pushing through accessibility challenges or coping with constant micro-aggressions. From this perspective, the call for reform is a call to equalise these energy costs across all members of society. The conservative impulse reflects the tendency to minimise energy expenditure by maintaining familiar systems, whereas the

reformist impulse reflects the recognition that current systems may impose unfair energy burdens on certain groups. Both are grounded in the same underlying concern with energy allocation, but they focus on different aspects of the problem.

Our account operates simultaneously at biological, psychological, social and moral levels, showing how they are related without reducing any level to the others. At the biological level, we have energy conservation mechanisms and interoceptive

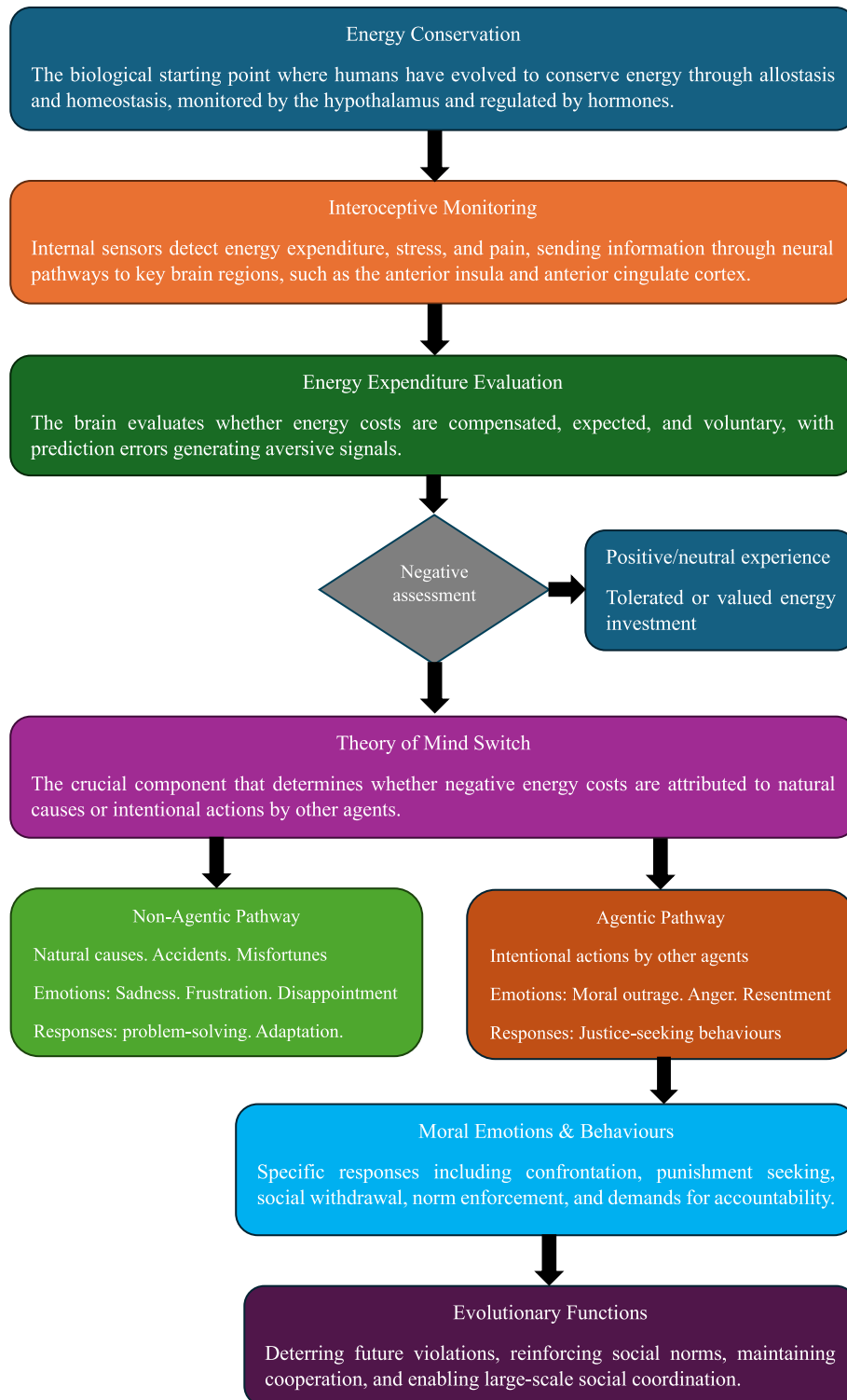


FIGURE 1 | The embodied cognitive mechanism of justice perception.

monitoring systems that evolved to help organisms manage their metabolic resources. At the psychological level, we have emotional responses and cognitive processes that evaluate social situations in terms of energy costs and benefits. At the social level, we have norms and institutions that regulate cooperation and conflict. At the moral level, we have principles of justice that guide individual and collective action (Figure 1).

6.3 | Merits, Limitations and Future Directions

The account presented here offers a way of linking several phenomena typically associated with justice to underlying physiological and motivational processes. It suggests that the physical discomfort often reported in response to unfair treatment may reflect embodied mechanisms for detecting uncompensated energetic demands. It also helps make sense of why people are frequently willing to incur personal costs to rectify perceived injustices. Although patterns of justice concern appear cross-culturally widespread, our framework accommodates variation in their specific norms and practices by treating them as context-sensitive calibrations of a shared underlying system. By tracing normative intuitions to evolved psychological mechanisms, this account may also contribute to clarifying the relation between descriptive and normative dimensions of moral life.

However, the energy-based approach has several limitations. First, although the theory provides an account of the affective basis of justice perception, it does not, by itself, determine the specific *normative content* of the cognitive judgements. The principles that define what counts as a 'just' or 'unjust' imposition of costs are heavily shaped by culture, social learning and explicit reasoning. Second, it is currently unclear how the energy costs of individuals or groups can be precisely estimated and quantified; without such input, applying the framework remains challenging. Third, some aspects of justice may involve considerations that go beyond energy allocation, such as symbolic recognition, dignity or respect, which cannot be easily reduced to energy terms. Fourth, the theory may struggle with cases where justice requires actions that are energetically costly but morally necessary, such as truth-telling that harms relationships. Finally, the framework may not fully capture the complexity of distributive justice, particularly in cases involving competing principles of need, merit and equality.

These limitations suggest several directions for future research. First, empirical studies could test specific predictions of the energy-based account, such as whether physiological measures of energy expenditure correlate with judgements of injustice. Second, philosophical work could explore how energy considerations interact with other moral values and principles. Third, cross-cultural research could examine how universal energy-based mechanisms interact with culturally specific justice norms.

7 | Conclusion

This paper has presented a physiological account of justice perception that grounds moral experience in embodied cognition and evolutionary history. By showing how energy conservation mechanisms give rise to justice concerns, we provide a

naturalistic foundation for moral intuitions while preserving their motivational force and normative significance. Our account integrates insights from biology, psychology and philosophy to offer a framework for understanding how humans perceive and respond to injustice. The proposed two-part model explains how ontologically objective features of the world are translated, via an ontologically subjective affective experience, into epistemically objective moral claims. Although the theory has limitations and requires further development, it provides a framework for understanding one of the most fundamental aspects of human moral experience.

The approach suggests two important conclusions about the relationship between scientific and moral understanding. First, scientific explanations of moral phenomena need not be reductionist. Our account shows how moral experiences can be both genuine features of human life and products of natural physiological processes. The fact that justice concerns have biological foundations does not make them less real or less important; it helps explain why they are so central to human experience. Second, moral philosophy and empirical science can be mutually informative. Understanding the physiological mechanisms underlying justice perception can inform moral and political theory by clarifying what kinds of social arrangements are likely to be *experienced* as just or unjust. Conversely, the philosophical analysis of moral concepts can guide empirical research by identifying the phenomena that need to be explained and the conceptual distinctions that matter for understanding them. Ultimately, by grounding moral experience in the biological imperatives of energy regulation and the cognitive complexities of social appraisal, our framework offers a new way to understand one of the most fundamental aspects of human moral life and also provides a concrete, naturalistic foundation for the core insights of enactivist ethics.

Author Contributions

Both authors, Shervin Mirzaeighazi and Trond A. Tjøstheim, contributed equally in writing and developing the idea for the article and performing the literature search.

Acknowledgements

We would like to thank Peter Gärdenfors, Rachael Brown, Douglas Porpora, and two anonymous reviewers of the Journal for the Theory of Social Behaviour for their valuable comments. All mistakes are ours.

Funding

The authors have nothing to report.

Ethics Statement

The authors have nothing to report.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

Endnotes

¹ See Searle (1995).

² See Shapiro and Spaulding (2025).

³ Barrett (2017) makes the analogy more broadly to a *body budget* which is credited by eating, sleeping and other growth-oriented processes, but debited by stressful events like trying to escape danger, fighting or giving public presentations.

⁴ For an interesting account regarding the connection between these affective states and punitive behaviour, see Yang et al. 2024.

⁵ This is, however, consistent with the experimental work, showing that people prefer inequality in the distribution of benefits when it is based on merit (Marshall et al. 1999; Konow 2001).

References

- Balliet, D., L. B. Mulder, and P. A. M. Van Lange. 2011. "Reward, Punishment, and Cooperation: A Meta-Analysis." *Psychological Bulletin* 137, no. 4: 594–614. <https://doi.org/10.1037/a0023489>.
- Barrett, L. F. 2017. *How Emotions Are Made: The Secret Life of the Brain*. Pan Macmillan.
- Batson, C. D., C. L. Kennedy, L.-A. Nord, et al. 2007. "Anger at Unfairness: Is It Moral Outrage?" *European Journal of Social Psychology* 37, no. 6: 1272–1285. <https://doi.org/10.1002/ejsp.434>.
- Beaulieu, J. M., and R. R. Gainetdinov. 2011. "The Physiology, Signaling, and Pharmacology of Dopamine Receptors." *Pharmacological Reviews* 63, no. 1: 182–217. <https://doi.org/10.1124/pr.110.002642>.
- Beckes, L., and J. A. Coan. 2011. "Social Baseline Theory: The Role of Social Proximity in Emotion and Economy of Action." *Social and Personality Psychology Compass* 5, no. 12: 976–988. <https://doi.org/10.1111/j.1751-9004.2011.00400.x>.
- Bellucci, G., J. A. Camilleri, V. Iyengar, S. B. Eickhoff, and F. Krueger. 2020. "The Emerging Neuroscience of Social Punishment: Meta-Analytic Evidence." *Neuroscience & Biobehavioral Reviews* 113: 426–439. <https://doi.org/10.1016/j.neubiorev.2020.04.011>.
- Blader, S. L., and Y. R. Chen. 2012. "Differentiating the Effects of Status and Power: A Justice Perspective." *Journal of Personality and Social Psychology* 102, no. 5: 994–1014. <https://doi.org/10.1037/a0026651>.
- Boehm, C. 2012. *Moral Origins: The Evolution of Virtue, Altruism, and Shame*. Basic Books.
- Brosnan, S. F., and F. B. de Waal. 2003. "Monkeys Reject Unequal Pay." *Nature* 425, no. 6955: 297–299. <https://doi.org/10.1038/nature01963>.
- Bshary, R., and A. S. Grutter. 2005. "Punishment and Partner Switching Cause Cooperative Behaviour in a Cleaning Mutualism." *Biology Letters* 1, no. 3: 396–399. <https://doi.org/10.1098/rsbl.2005.0344>.
- Buckholtz, J. W., and R. Marois. 2012. "The Roots of Modern Justice: Cognitive and Neural Foundations of Social Norms and Their Enforcement." *Nature Neuroscience* 15, no. 5: 655–661. <https://doi.org/10.1038/nn.3087>.
- Cannon, W. B. 1932. *The Wisdom of the Body*. Kegan Paul and Co., Ltd.
- Carlsmith, K. M., J. M. Darley, and P. H. Robinson. 2002. "Why Do We Punish? Deterrence and Just Deserts as Motives for Punishment." *Journal of Personality and Social Psychology* 83, no. 2: 284–299. <https://doi.org/10.1037/0022-3514.83.2.284>.
- Carpenter, J., and D. Dolifka. 2017. "Exploitation Aversion: When Financial Incentives Fail to Motivate Agents." *Journal of Economic Psychology* 61: 213–224. <https://doi.org/10.1016/j.joep.2017.04.006>.
- Clark, A. 2016. *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. Oxford University Press.
- Cohen, G. A. 1989. "On the Currency of Egalitarian Justice." *Ethics* 99, no. 4: 906–944. <https://doi.org/10.1086/293126>.
- Craig, A. D. 2002. "How Do You Feel? Interoception: The Sense of the Physiological Condition of the Body." *Nature Reviews Neuroscience* 3, no. 8: 655–666. <https://doi.org/10.1038/nrn894>.
- Cushman, F. 2014. "Punishment in Humans: From Intuitions to Institutions." *Philosophy Compass* 10, no. 2: 117–133. <https://doi.org/10.1111/phc3.12192>.
- Decety, J., and K. J. Yoder. 2017. "The Emerging Social Neuroscience of Justice Motivation." *Trends in Cognitive Sciences* 21, no. 1: 6–14. <https://doi.org/10.1016/j.tics.2016.10.008>.
- De Cremer, D., and T. R. Tyler. 2005. "Managing Group Behaviour: The Interplay Between Fairness, Self, and Cooperation." In *Advances in Experimental Social Psychology*, edited by M. Zanna, 151–218. Academic.
- De Waal, F. B. M. 1996. *Good Natured: The Origins of Right and Wrong in Humans and Other Animals*. Harvard University Press.
- Dong, M., J. W. Van Prooijen, and P. A. van Lange. 2022. "Strategic Exploitation by Higher-Status People Incurs Harsher third-party Punishment." *Social Psychology* 53, no. 4: 209–220. <https://doi.org/10.1027/1864-9335/a000493>.
- Dubin, A. E., and A. Patapoutian. 2010. "Nociceptors: The Sensors of the Pain Pathway." *Journal of Clinical Investigation* 120, no. 11: 3760–3772. <https://doi.org/10.1172/jci42843>.
- Dworkin, R. 2000. *Sovereign Virtue: The Theory and Practice of Equality*. Harvard University Press.
- Ellard, J. H., A. Harvey, and M. J. Callan. 2016. "The Justice Motive." In *Handbook of Social Justice Theory and Research*, edited by C. Sabbagh and M. Schmitt, 127–143. Springer.
- Fast, N. J., N. Halevy, and A. D. Galinsky. 2012. "The Destructive Nature of Power Without Status." *Journal of Experimental Social Psychology* 48, no. 1: 391–394. <https://doi.org/10.1016/j.jesp.2011.07.013>.
- Fehr, E., and S. Gächter. 2002. "Altruistic Punishment in Humans." *Nature* 415, no. 6868: 137–140. <https://doi.org/10.1038/415137a>.
- Fiske, S. T., C. H. Dupree, G. Nicolas, and J. K. Swencionis. 2016. "Status, Power, and Intergroup Relations: The Personal Is the Societal." *Current Opinion in Psychology* 11: 44–48. <https://doi.org/10.1016/j.copsy.2016.05.012>.
- Gallagher, S. 2024. "Reenacting Ethics: Rejoindre l'aventure de Varela." (Translated by P. Lorelle). In *Francisco Varela: Une Pensée Actuelle*, edited by N. Depraz, 65–78. Hermann.
- Geraci, A. 2021. "Toddlers' Expectations of Corporal Third-Party Punishments Against the Non-Defender Puppet." *Journal of Experimental Child Psychology* 210: 105199. <https://doi.org/10.1016/j.jecp.2021.105199>.
- Gintis, H., J. Henrich, S. Bowles, R. Boyd, and E. Fehr. 2008. "Strong Reciprocity and the Roots of Human Morality." *Social Justice Research* 21, no. 2: 241–253. <https://doi.org/10.1007/s11211-008-0067-y>.
- Goldinger, S. D., H. M. Kleider, T. Azuma, and D. R. Beike. 2003. "Blaming the Victim Under Memory Load." *Psychological Science* 14, no. 1: 81–85. <https://doi.org/10.1111/1467-9280.01423>.
- Gollwitzer, M., and J. W. van Prooijen. 2016. "Psychology of Justice." In *Handbook of Social Justice Theory and Research*, edited by C. Sabbagh and M. Schmitt, 61–82. Springer.
- Gosnell, B. A., and A. S. Levine. 2009. "Reward Systems and Food Intake: Role of Opioids." Supplement, *International Journal of Obesity* 33, no. S2: S54–S58. <https://doi.org/10.1038/ijo.2009.73>.
- Haidt, J. 2003. "The Moral Emotions." In *Handbook of Affective Sciences*, edited by R. J. Davidson, K. R. Scherer, and H. H. Goldsmith, 852–870. Oxford University Press.
- Haidt, J. 2007. "The New Synthesis in Moral Psychology." *Science* 316, no. 5827: 998–1002. <https://doi.org/10.1126/science.1137651>.
- Haidt, J. 2012. *The Righteous Mind: Why Good People Are Divided by Politics and Religion*. Penguin.

- Hamilton, W. D. 1964a. "The Genetical Evolution of Social Behaviour. I." *Journal of Theoretical Biology* 7, no. 1: 1–16. [https://doi.org/10.1016/0022-5193\(64\)90038-4](https://doi.org/10.1016/0022-5193(64)90038-4).
- Hamilton, W. D. 1964b. "The Genetical Evolution of Social Behaviour. II." *Journal of Theoretical Biology* 7, no. 1: 17–52. [https://doi.org/10.1016/0022-5193\(64\)90039-6](https://doi.org/10.1016/0022-5193(64)90039-6).
- Hamlin, J. K. 2013. "Moral Judgment and Action in Preverbal Infants and Toddlers: Evidence for an Innate Moral Sense." *Current Directions in Psychological Science* 22, no. 3: 186–193. <https://doi.org/10.1177/0963721412470687>.
- Hamlin, J. K., K. Wynn, P. Bloom, and N. Mahajan. 2011. "How Infants and Toddlers React to Antisocial Others." *Proceedings of the National Academy of Sciences* 108, no. 50: 19931–19936. <https://doi.org/10.1073/pnas.1110306108>.
- Haselton, M., and D. Buss. 2009. "Error Management Theory and the Evolution of Misbeliefs." *Behavioral and Brain Sciences* 32, no. 6: 522–523. <https://doi.org/10.1017/s0140525x09991440>.
- Hoffman, M. B., and T. H. Goldsmith. 2004. "The Biological Roots of Punishment." *Ohio State Journal of Criminal Law* 1: 627. <https://gruterinstitute.org/the-biological-roots-of-punishment/>.
- Hohwy, J. 2013. *The Predictive Mind*. Oxford University Press.
- Hume, D. 1975. "An Enquiry Concerning the Principles of Morals." In *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*, edited by L. A. Selby-Bigge and P. H. Nidditch, Rev. ed. Clarendon Press.
- Hutto, D. D. 2025. "Enacting Phronesis: Some Deliberations About Enactive Ethics." *Mind & Society* 24, no. 24: 7–35. <https://doi.org/10.1007/s11299-025-00316-1>.
- Jacob, F. 1977. "Evolution and Tinkering." *Science* 196, no. 4295: 1161–1166. <https://doi.org/10.1126/science.860134>.
- Jordan, J. J., M. Hoffman, P. Bloom, and D. G. Rand. 2016. "Third-Party Punishment as a Costly Signal of Trustworthiness." *Nature* 530, no. 7591: 473–476. <https://doi.org/10.1038/nature16981>.
- Kandel, E. R., J. Koester, S. Mack, and S. Siegelbaum. 2021. *Principles of Neural Science, 6e, AccessBiomedical Science*. McGraw Hill Med.
- Kiverstein, J., M. D. Kirchhoff, and M. Thacker. 2022. "An Embodied Predictive Processing Theory of Pain Experience." *Review of Philosophy and Psychology* 13, no. 4: 973–998. <https://doi.org/10.1007/s13164-022-00616-2>.
- Konow, J. 2001. "Fair and Square: The Four Sides of Distributive Justice." *Journal of Economic Behavior & Organization* 46, no. 2: 137–164. [https://doi.org/10.1016/s0167-2681\(01\)00194-9](https://doi.org/10.1016/s0167-2681(01)00194-9).
- Kringelbach, M. L. 2010. "The Hedonic Brain: A Functional Neuroanatomy of Human Pleasure." In *Pleasures of the Brain*, edited by M. L. Kringelbach and K. C. Berridge, 202–221. Oxford University Press.
- Ledoux, J. 2021. "As Soon as There Was Life, There Was Danger: The Deep History of Survival Behaviours and the Shallower History of Consciousness." *Philosophical Transactions of the Royal Society B: Biological Sciences* 377, no. 1844: 20210292. <https://doi.org/10.1098/rstb.2021.0292>.
- Le Merrer, J., J. A. Becker, K. Befort, and B. L. Kieffer. 2009. "Reward Processing by the Opioid System in the Brain." *Physiological Reviews* 89, no. 4: 1379–1412. <https://doi.org/10.1152/physrev.00005.2009>.
- MacCormack, J. K., and K. A. Muscatell. 2019. "The Metabolic Mind: A Role for Leptin and Ghrelin in Affect and Social Cognition." *Social and Personality Psychology Compass* 13, no. 9: e12496. <https://doi.org/10.1111/spc3.12496>.
- Manrique, H., H. Zeidler, G. Roberts, et al. 2021. "The Psychological Foundations of Reputation-Based Cooperation." *Philosophical Transactions of the Royal Society B* 376, no. 1838: 20200287. <https://doi.org/10.1098/rstb.2020.0287>.
- Marlowe, F. W. 2009. "Hadza Cooperation: Second-Party Punishment, Yes; Third-Party Punishment." *Human Nature* 20, no. 4: 417–430. <https://doi.org/10.1007/s12110-009-9072-6>.
- Marlowe, F. W., J. C. Berbesque, C. Barrett, A. Bolyanatz, M. Gurven, and D. Tracer. 2011. "The 'Spiteful' Origins of Human Cooperation." *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 278, no. 1715: 2159–2164. <https://doi.org/10.1098/rspb.2010.2342>.
- Marshall, G., A. Swift, D. Routh, and C. Burgoyne. 1999. "What Is and What Ought To Be: Popular Beliefs About Distributive Justice in Thirteen Countries." *European Sociological Review* 15, no. 4: 349–367. <https://doi.org/10.1093/oxfordjournals.esr.a018270>.
- Menzies, J. R. W., K. P. Skibicka, G. Leng, and S. L. Dickson. 2013. "Ghrelin, Reward and Motivation." *Endocrine Development* 25: 101–111. <https://doi.org/10.1159/000346058>.
- Nagel, T. 2005. "The Problem of Global Justice." *Philosophy and Public Affairs* 33, no. 2: 113–147. <https://doi.org/10.1111/j.1088-4963.2005.00027.x>.
- Peters, A., B. S. McEwen, and K. Friston. 2017. "Uncertainty and Stress: Why It Causes Diseases and How It Is Mastered by the Brain." *Progress in Neurobiology* 156: 164–188. <https://doi.org/10.1016/j.pneurobio.2017.05.004>.
- Raphael, D. D. 2001. *Concepts of Justice*. Clarendon Press.
- Rawls, J. 1971. *A Theory of Justice*. Harvard University Press.
- Ruff, C. C., and E. Fehr. 2014. "The Neurobiology of Rewards and Values in Social Decision Making." *Nature Reviews Neuroscience* 15, no. 8: 549–562. <https://doi.org/10.1038/nrn3776>.
- Searle, J. R. 1995. *The Construction of Social Reality*. Free Press.
- Sen, A. 1980. "Equality of What?" In *The Tanner Lectures on Human Values*, edited by S. McMurrin, Vol. 1, 195–220. Cambridge University Press.
- Shapiro, L., and S. Spaulding. 2025. "Embodied Cognition." In *The Stanford Encyclopedia of Philosophy*, edited by E. N. Zalta and U. Nodelman. <https://plato.stanford.edu/archives/sum2025/entries/embodyed-cognition/>.
- Sidgwick, H. 1907. *The Methods of Ethics*. (Original work published 1874). Macmillan.
- Solomon, R. C. 1988. "On Emotions as Judgments." *American Philosophical Quarterly* 25, no. 2: 183–191. <https://philpapers.org/rec/SOLOEA>.
- Sterling, P. 2004. "Principles of Allostasis: Optimal Design, Predictive Regulation, Pathophysiology." *Physiology & Behavior* 82, no. 1: 1–15. <https://doi.org/10.1017/CBO9781316257081.004>.
- Takhvar, M. 1988. "Play and Theories of Play: A Review of the Literature." *Early Child Development and Care* 39, no. 1: 221–244. <https://doi.org/10.1080/0300443880390117>.
- Thomas, A. L., S. M. Davis, and H. A. Dierick. 2015. "Of Fighting Flies, Mice, and Men: Are Some of the Molecular and Neuronal Mechanisms of Aggression Universal in the Animal Kingdom?" *PLoS Genetics* 11, no. 8: e1005416. <https://doi.org/10.1371/journal.pgen.1005416>.
- Trivers, R. L. 1971. "The Evolution of Reciprocal Altruism." *Quarterly Review of Biology* 46, no. 1: 35–57. <https://doi.org/10.1086/406755>.
- Tyler, T. R., and S. L. Blader. 2000. *Cooperation in Groups: Procedural Justice, Social Identity, and Behavioral Engagement*. Taylor & Francis.
- Tyler, T. R., and E. A. Lind. 1992. "A Relational Model of Authority in Groups." In *Advances in Experimental Social Psychology*, edited by M. Zanna, Vol. 25, 115–191. Academic Press. [https://doi.org/10.1016/s0065-2601\(08\)60283-x](https://doi.org/10.1016/s0065-2601(08)60283-x).
- Van den Bos, K., and E. A. Lind. 2002. "Uncertainty Management by Means of Fairness Judgments." In *Advances in Experimental Social*

Psychology, edited by M. P. Zanna, Vol. 34, 1–60. Academic. [https://doi.org/10.1016/s0065-2601\(02\)80003-x](https://doi.org/10.1016/s0065-2601(02)80003-x).

Van Vugt, M., R. Hogan, and R. B. Kaiser. 2008. “Leadership, Followership, and Evolution.” *American Psychologist* 63, no. 3: 182–196. <https://doi.org/10.1037/0003-066X.63.3.182>.

Waxman, S. G., T. R. Cummins, S. Dib-Hajj, J. Fjell, and J. A. Black. 1999. “Sodium Channels, Excitability of Primary Sensory Neurons, and the Molecular Basis of Pain.” *Muscle & Nerve* 22, no. 9: 1177–1187. [https://doi.org/10.1002/\(sici\)1097-4598\(199909\)22:9<1177::aid-mus3>3.3.co;2-g](https://doi.org/10.1002/(sici)1097-4598(199909)22:9<1177::aid-mus3>3.3.co;2-g).

Wenseleers, T., and F. L. W. Ratnieks. 2006. “Enforced Altruism in Insect Societies.” *Nature* 444, no. 7115: 50. <https://doi.org/10.1038/444050a>.

Yang, Q., M. Hoffman, and F. Krueger. 2024. “The Science of Justice: The Neuropsychology of Social Punishment.” *Neuroscience & Biobehavioral Reviews* 157: 105525. <https://doi.org/10.1016/j.neubiorev.2023.105525>.

Young, L., F. Cushman, M. Hauser, and R. Saxe. 2007. “The Neural Basis of the Interaction Between Theory of Mind and Moral Judgment.” *Proceedings of the National Academy of Sciences* 104, no. 20: 8235–8240. <https://doi.org/10.1073/pnas.0701408104>.