European Journal of Social Psychology

WILEY
EASP

OUTLOOK

The Psychology of State Punishment

Jordan Wylie[1] | Connie P. Y. Chiu[2] | Nicolette M. Dakin[2] | William Cunningham[3] | Ana Gantman[2,4]

[1]Department of Psychology & Neuroscience, Boston, College, Chestnut Hill, Massachusetts, USA | [2]Department of Psychology, The Graduate Center, CUNY, New York City, New York, USA | [3]Department of Psychology, University of Toronto, Toronto, Canada | [4]Department of Psychology, Brooklyn College, Brooklyn, New York, USA

Correspondence: Ana Gantman (Ana.Gantman@brooklyn.cuny.edu)

Received: 8 October 2024 | **Accepted:** 13 January 2025

Keywords: institutions | moral psychology | state punishment | third-party punishment

ABSTRACT

A significant amount of punishment that happens in society is state punishment, that is, third-party punishment carried out by an organized political community in response to a rule violation. We argue that a complete psychology of punishment must consider state punishment as a distinct form. State punishment is a unique type of punishment because it is a special case of third-party punishment, pre-specified to occur after the violation of official rules and policies, carried out by people acting on behalf of a nation or government. State punishment, especially as compared to interpersonal punishment, is regarded as a legitimate form of violence, which communicates not just disapproval but information about procedures and power. Moreover, state punishment is made possible by state rules, which, unlike norms, are formalized, can be fully articulated and are perfectly transmissible across generations. We end the paper with implications for the psychology of punishment more broadly and future directions for better understanding the unique psychology of state punishment.

1 | Introduction

There are many rules to follow. For example, to board an airplane in the United States in the year 2024, following are a few things that one must do by law: put 3.4 oz of liquids or less in individual containers, empty pockets, allow screening of bags, take off shoes and walk through metal detector or whole-body scan. One might also do things in the airport like speak quietly and wait patiently in line. The former are the official Transportation Security Authority (TSA) rules, and the latter are the social norms governing how people behave in the screening line. State punishment is what follows breaking the official, explicit and codified rules (and getting caught). Official government rules, like the TSA regulations for boarding an airplane, are meant to follow the rule of law, meaning they are general (i.e., not targeting specific individuals or situations), public, non-retroactive, consistent, clear, possible to follow, stable and applying also to those who enforce them (Fuller 1964). Indeed, cross-culturally, people expect laws to follow these principles (Hannikainen et al. 2021). Rules of this kind are enforced by the third-party institutions that put the rule in place, and they tend to come with a range of pre-specified punishments.

If the state—here meaning a nation or territory considered as an organized political community under one government—put the rule in place, then one can expect the state to be the one to punish those who break it. This kind of state-sponsored punishment is a form of third-party punishment or punishment administered by someone not directly affected by the transgression (Fehr and Fischbacher 2004; Buckholtz et al. 2008; Jensen, Call, and Tomasello 2007; Riedl et al. 2012). Third-party punishment has been extensively theorized as the engine behind uniquely human, large-scale cooperation, allowing people to cooperate with non-kin and to punish free riding (for a review, see Wylie and Gantman 2024). Accordingly, third-party punishment is also the key to understanding human institutions like nations, cities

and universities. It scaffolds the formation of institutions (a centralized purveyor of third-party punishment) and—when studied in the context of economic games—reflects the institutions that players are a part of (Henrich and Muthukrishna 2021).

Consider state punishment in contrast with interpersonal punishment. For example, if a person interferes with the screening process administered by the TSA, they can face up to $14,950 in fines owed to the US government (Transportation Security Administration 2022). If someone were loudly talking or wearing a ball gown in the screening line they might overhear some comments about them or even get into a confrontation. These are interpersonal punishments. Note that if a person were wearing a ballgown in the screening line, and they received a fine from the TSA they would rightfully contest it. This behaviour breaks a social norm but not a law.

With this paper, we seek to better understand the psychology of punishment by focusing specifically on state punishment. State punishment is unique because it is a specific form of third-party punishment, pre-specified to occur after the violation of official rules and policies and carried out by people acting on behalf of a nation or government and not by the directly affected party. State punishment reveals novel insights about the psychology of punishment more generally and, in particular, reveals how everyday people and agents of the state enact punishment.

1.1 | State Punishment Solves Coordination Problems

State-sponsored punishment, like a fine for failing to comply with TSA screening procedures and all forms of third-party punishment, can be said to solve significant coordination problems that humans living in large groups face (Hadfield and Weingast 2012). Modern humans live in groups with both kin and non-kin and must share resources, balancing the needs of individuals with those of the group (see e.g., Ostrom 1990). Third-party punishment promotes cooperation and coordination (e.g., Fehr and Gächter 2000, 2002; Fehr, Fischbacher, and Gächter 2002; see Wylie and Gantman 2024 for review), sanctioning free riders who benefit from group life without contributing, ensuring no individual takes more than their fair share and, arguably, making possible the proliferation and flourishing of large-scale societies (e.g., Henrich and Muthukrishna 2021; Lie-Panis et al. 2023).

Critically, this form of third-party punishment is administered by and, as a function, is legitimized by the state (and its associated institutions). This reduces the burden of punishment of any one individual, and it codifies behaviours that often directly contribute to social order and functioning (e.g., Fehr and Fischbacher 2004), though not always for the better (e.g., Wylie and Gantman 2024). Thus, by codifying norms into rules and laws, the state makes the boundaries of many behaviours, and the corresponding punishments for transgressing them, significantly clearer and easier to comply with. In turn, this makes the behaviour of the people governed by these rules easier to predict. The rules are designed to align people's behaviour with the desires of the state (Scott 2020), often compelling actions individuals might not otherwise take by conveying the potential for punishment in cases of non-compliance. In this way, state rules and state punishment play a large role in both creating and maintaining social order in today's complex societies (Henrich and Muthukrishna 2021). Further, institutions are also often expected to do this in ways that are fair (e.g., Tyler 1997), and when they do, they are considered more legitimate, and better able to induce compliance (Sunshine and Tyler 2003). Such institutions may be the explanation for the particular success of humans, a kind of codified cooperation (Henrich and Muthukrishna 2021; Lie-Panis et al. 2023).

Institutional punishment (e.g., state punishment) also specifically solves the second-order free rider problem; it benefits the group to incur costs to punish free riders and ensures that those who benefit from the group also contribute to it (i.e., solves the first-order free rider problem). But people can also free ride on the benefits of costly third-party punishment without engaging in it themselves (i.e., the second order free rider problem; e.g., Boyd and Richerson 1992; Ozono et al. 2016). One way to solve this problem (aside from socially rewarding those willing to engage in third-party punishment; Gintis et al. 2001; Jordan et al. 2016) is to consolidate the third-party punishment to be administered into institutions (Baldassarri and Grossman 2011). The use of state punishment as third-party punishment also wards off a cycle of vengeance that second-party punishment can ignite (see Raihani and Bshary 2019 for review). Note that sometimes it may be that the state engages in second-party punishment, for instance, when the crime itself is against the state. In minor cases, this might be when police enforce rules like making sure people have required state IDs (as they often do; Graeber 2015), and more extreme cases include the prosecution of treason where, notably, punishments for treason often include death.

In sum, institutions (like the state) are understood as the largest-scale form of cooperation that humans have devised, having first evolved to favour kin, engage in reciprocity and cooperate based on reputation or group membership (Muthukrishna et al. 2017). Critically though, these forms of cooperation did not successively replace each other but instead co-exist. This makes possible new forms of 'cooperation', for example, an individual engaging in reciprocity with another in an institutional role (i.e., bribery), which can benefit the individual but undermine cooperation at the institutional level (Muthukrishna et al. 2017). State punishments also solve complex social coordination problems by domination, in which coordination is achieved by oppressing some and benefiting others (Wylie and Gantman 2024).

1.2 | State Punishment Is a Unique Form of Third-Party Punishment

1.2.1 | Legitimate Use of Violence for Enforcement

State punishment is a special form of third-party punishment. State punishment follows violations of state law and policy. Only once an official rule has been deemed to be broken do these (sometimes very harmful) sanctions become perceived as both possible and legitimate. State punishment is also the locus of a significant amount of punishment in people's lives. It includes some of the harshest punishments possible (see e.g., Bernhard, Cushman, and LeBaron 2022). For example, one form of state punishment is the death penalty, which a majority of US Americans support in the case of murder (Pew Research

Center 2021), for largely retributive reasons (Griffin 2021). And research suggests that a great deal of punishment is motivated not by the desire to prevent future rule breaking but to inflict proportionate harm onto those who have violated the rules (Carlsmith, Darley, and Robinson 2002). Moreover, the use of violent force by the state is seen as legitimate. Indeed, some have suggested that claims to legitimate violence are integral to the very definition of what makes a state. For example, Weber (1919) suggested that the defining characteristic of the state is the monopoly on the legitimate use of violence. Individuals may have some capacity to exercise physical force, but these powers are ultimately circumscribed by the state (e.g., in the United States, a parent—another significant source of punishment in people's lives—may spank their child to discipline them but the state may interfere if they exert too much force). State punishment has a unique claim to this legitimate use of force. That is, in contrast to theories of punishment that primarily focus on promoting cooperation, third-party punishment, in the form of state punishment, uniquely wields the power to inflict both psychological and physical violence (Foucault 1977) on the people in its purview.

The legitimate use of violence is critical to understanding third-party punishment in the real world. People are incredibly sensitive to which rules can be broken and when (Levine et al. 2024) and will go to great lengths to avoid being outright disobedient (Qian et al. 2024), but when someone does break a law, there is an available avenue for recourse. When the process of legal rule enforcement is working correctly (i.e., is fair and legitimate), this cause-and-effect relationship serves to restore social order and appropriately redress harms. Take, for example, the case of drunk driving. If someone drives under the influence and you notice and alert the police, the law against that behaviour means that the police can take that person off the road and disincentivize them from doing it again. The perceived legitimacy of punishments that follow breaking laws makes them powerful, even for the everyday person. And a great deal of work in psychology has suggested that legitimacy and related concepts like procedural fairness (i.e., judgements that legal authorities act with fairness in the procedures associated with their decisions; Tyler 1997) are important elements in whether people trust in the state and are likely to comply with rules (Tyler 2003). That is, central to the power of state punishments is the perception that said punishment and authorities are legitimate.

When states and their punishments achieve perceived legitimacy, they also afford everyday people a legitimate (albeit indirect) route to punish others. This is true even when the rule violations are not motivated by reasons that reasonable people would agree are justifiable for police involvement. For example, if someone were to occupy a space that others did not approve of (i.e., someone was loitering), the law against loitering means that those people can call the police to remove that person. Sometimes this can be wielded to enforce property rights—a function essential for the stability and enforcement of social contracts (and central to state formation; see e.g., Locke 1689/1988). Other times, it can be used to exert social control or reinforce biases, leading to disproportionate enforcement against some people more than others. That is, the punishment of state rules extends beyond simply addressing non-compliance to maintain social order. Codified rules afford punishment of violations of both their letter (i.e., the literal text) and of their spirit (i.e., the understood intention behind the rule)—and potentially to flexibly select between them. For example, in the case of loitering, one may be punished for literally occupying public or private property or for violating the perceived intentions underlying the rule, such as appearing out of place in a particular context. Critically, both paths lead to lay judgements of culpability (Garcia, Chen, and Gordon 2014; see also Struchiner, Hannikainen, and de Almeida 2020), and so enforcement of loitering can legitimize the use of punishment even in cases where the desire for punishment stems from 'pretextual' reasons.

In this way, the legitimacy of the law not only empowers state authorities to administer punishment but also provides citizens with an avenue to enforce rules and punish individuals indirectly. When someone calls the police on the loiterer, they are not just seeking the removal of an unwanted presence, they are leveraging the legitimacy of legal authority to carry out their personal desires. When citizens opt to do this, they are able to carry out not only their personal desires but do so with a pre-existing justification—the law proscribes some behaviour of the person to be punished, though it may not be the behaviour that evoked the desire to punish them (Wylie and Gantman 2023, 2024).

1.2.2 | Unique Communication

State punishment, in addition to its monopoly over legitimate violence, also communicates distinct information compared to other forms of punishment. This is, in part, because state punishment requires that individuals dole out punishment not as their individual selves alone but within their role as a representative of that state. People who are actors within the state or employed by it are part of a hierarchy where people are sorted into highly specialized roles (Graeber 2015). These roles circumscribe what behaviour is possible for that person to do within the state (e.g., legislators make laws, police enforce them, etc.). Typically, when we seek to understand others' behaviour, we engage in theory of mind (e.g., Saxe and Kanwisher 2013) or take an intentional stance (Dennett 1988/1993), trying to infer others' beliefs, desires and intentions. We then typically use these mental state attributions to determine responsibility, blame and punishment (see e.g., Malle, Guglielmo, and Monroe 2014)—or lessen them when those mental state attributions do not apply (e.g., not blaming someone for speeding because of a miscalibrated speedometer; Turri 2019). However, when a person is acting within their role in the state, we are more likely to make meaning of their actions by taking an institutional stance, meaning that we understand that their behaviour is explicable from their role within the institution rather than their particular beliefs and desires as an individual (Jara-Ettinger and Dunham 2024). This is well-illustrated by the need to hide the identity of the people who carry out state executions. We are not meant to infer malice, remorse, the will of justice or any other mental state in the executioner. We are meant to only see them in their role. And indeed, if this same person killed someone while acting outside of this role, they would be subject to state punishment themselves. This is critical because punishment in general, and state punishment in particular, are often theorized to be about communication,

moral education or expression (rather than just disincentive; e.g., Sarin et al. 2021; McAdams 2015; Kahan 1996; Sunstein 1996; Hampton 1984), specifically of the values of a system, state or community (albeit imperfectly), which provide meaning and express which behaviours are disliked. State punishment communicates something additional or something else entirely—that an official law was broken, that others should not break this law and that extensive legal processes were carried out to ensure the use of violence to enforce the law was legitimate.

Moreover, when a person acts as a representative of the state, it is difficult to decide how much blame, responsibility or punishment they deserve if they commit moral violations within it. Typically, people who engage in costly third-party punishment reap reputational rewards (Jordan et al. 2016). But individuals who carry out state punishment do not stand to reap those same benefits, likely because we are not warranted in inferring their motivation for engaging in punishment (Jordan and Kteily 2023); their motive is their job description. Moreover, people are more likely to hold responsible people who make decisions that end in harm than those who carry out their directives. Indeed, while judgements of intent for typical agents tend to be sensitive to moral concern (Knobe 2010), the intent of those who carry out the orders of others are not (Gantman et al. 2020).[1] The violence that the state can enact seems not to have an intentional author in quite the same way as other forms of punishment.

1.2.3 | Perfectly Transmissible

Legal codes are also passed down with perfect fidelity. They are explicitly written down, made public and often require procedures to change—a feature identified as critical for cumulative cultural knowledge (Henrich and Gil-White 2001). Unlike social norms, which can have imperfect transmission and can change or even be abandoned over time (e.g., Centola et al. 2018; see Gelfand, Gavrilets, and Nunn 2024 for review of norm dynamics), laws provide a means to formalize and preserve certain behaviours and corresponding punishments, ensuring that they are maintained and enforced even as individuals, governments and societies come and go. This crystallization helps to stabilize norms and protect against the erosion of critical cultural knowledge (Bicchieri et al. 2023; Keizer, Lindenberg, and Steg 2008; Legros and Cislaghi 2020), and it also allows agents of the state (e.g., judges) to act in ways that are consistent across a wide range of contexts, bolstering the legitimacy of the rule of law (Hannikainen et al. 2022). Laws are also hard to change, for example, the Jones Act of 1920 currently influences how cargo is transported by sea in the United States (see Wylie and Gantman 2024). Laws can be preserved, if not observed, for centuries; one can still view the Code of Hammurabi (and, in theory, read it), which details the legal code of Ancient Babylon etched in Basalt during the reign of King Hammurabi (1792–1750 BCE).

The transmission of rules also allows for planning and predictability. Living with other people, and attempting to cooperate and coordinate with them, requires some expertise in predicting what they will do (see also Hadfield and Weingast 2012). Rules, like other articulated sets of behavioural expectations, significantly help to reduce prediction errors—people in a given society often assume that others will follow the rule (especially within tight cultures; see Gelfand, Gavrilets, and Nunn 2024), and the rule articulates how to and how not to behave (Wheeler et al. 2020). Driving provides a useful example—because of the rules, people in, for example, the United Kingdom expect (usually correctly) people to drive on the left side of the road while those in the United States expect them to drive on the right. Rules help constrain the possible actions available in a given context. Codified rules then are both perfectly transmissible and promoters of human predictability.

1.2.4 | Fully Articulable

Rules are also uniquely articulable. While norms can remain unsaid, subjective and context-dependent, official rules are required to be explicit, subject to some objective standards and be articulable (albeit not always clear; Andrews 2009). Indeed, in the United States, limits are placed on how vague a law can be through the void-for-vagueness doctrine. Laws that are so vague as to render them impossible for an ordinary person to follow are deemed unconstitutional (see Mannheimer 2019)[2]. Moreover, rules can differ in how thick or thin they are, requiring more or less experience, respectively, to properly interpret whether they apply (Datson 2022). Within this range, there are explicit guidelines that people use to determine whether the rule was followed or if its spirit or letter was violated (e.g., de Almeida, Struchiner, and Hannikainen 2023), and they tend to defer to the letter when the need for coordination is made salient (Hannikainen et al. 2022). People can and do debate when rules should apply, but it is this fully articulated nature that makes such discussions possible. Even if the letter of the law does not perfectly delineate the forbidden behaviours exactly, such as what constitutes 'obscenity' under the law, it provides a standard through which various behaviours can be judged[3] and even through which people can find a space between following and not following the rule (i.e., a loophole; Qian et al. 2024).

Consider these processes and features of legal rules in contrast to social norms which can, and often do, remain unsaid. For example, if one person gets too close to another in conversation, the other person may step back. If asked why, they might be able to explain that their conversation partner violated their personal space but not be able to say exactly how close is too close (Andrews 2009; Cooperrider 2022). But, if there is a rule about how far apart people must stand, say, waiting behind a painted-on line at the grocery store or Customs and Border Control at the airport, the correct distance between two parties is perfectly articulable. Further, one might receive conflicting information about how close one can get in conversation with other people waiting in the customs line, but the appropriate distance from the customs officer is made very clear.

The special articulability of codified rules has implications for what kinds of violations people opt to punish. If someone has broken the letter of the law, but they are not perceived as doing anything particularly wrong or harmful—say because the thing they did is something people do frequently, like jaywalking in New York City—people may not judge the punishment to be warranted at all (Wylie and Gantman, 2023). Indeed, people may

only want to see that rule enforced if they want to punish the person for some other, pretextual reason (Wylie and Gantman 2023, 2024; Wylie et al. 2024). In these cases, the fully articulable nature of rules (vs. social norms) becomes significant. If someone really wants to punish someone for breaking a social norm, or being a jerk, they can actually point to the violation of a law like jaywalking, which was clearly broken according to its most literal articulation and interpretation, to make state punishment a viable option for carrying out one's personal desire for punishment (Wylie and Gantman 2023, 2024; Wylie et al. 2024).

2 | Implications

In sum, state punishment is a distinct form of third-party punishment that can help us better understand punishment conceptually and also punishment as-lived. State punishment holds a uniquely legitimate claim to the use of violence, communicates not only societal values but also the rule of law itself, and does not confer similar levels of reputational gain as other forms of costly punishment because it is carried out independently of the motives of individuals. State punishment is the enforcement of state rules that are both perfectly transmissible and fully articulable, two features which give punishment particular claims to legitimacy, while at the same time, allowing for its motivated use.

2.1 | Unidealizing State Punishment

No theory of punishment is complete without a consideration of how punishment really functions. Here, we suggest that state punishment, a significant source of punishment in people's lives, is unique and gives us an important window into the psychology of punishment more broadly. When researchers examine punishment, they must consider not just what punishment might offer in theory but also account for the practical realities and injustices present in the world (Wylie and Gantman 2024; Mills 2005). State punishment does not always live up to the ideals of the rule of law—the laws being consistent, applying equally to all in their jurisdiction, public, non-retroactive, comprehensible and more (Fuller 1964; Hannikainen et al. 2021). Indeed, instead of the rule of law, some states either partly or completely enact the rule by law, in which people acting on behalf of the state can dominate other citizens. From the point of view of history, this more retributive and violent use of state punishment is obvious—many laws and their enforcement have led to some of the most ruthless acts of violence in history (e.g., the Virginia Slave Codes of 1705 and the Indian Removal Act of 1830 in the United States; the Final Solution of 1941–1945 in Nazi Germany). Without centring state punishment in our theorizing, researchers run the risk of overemphasizing the functional role of punishment in cooperation and under-theorizing about its role in domination (Wylie and Gantman 2024).

2.2 | Future Directions

2.2.1 | On the Uniqueness of State Punishment

State punishments meaningfully lower the potential costs associated with third-party punishments. A third-party who calls the police to intervene in an altercation no longer runs the risk of themselves being physically harmed (they do not have to physically approach the situation). Future research should examine when people prefer to opt to enforce official rules, or delegate to state officials for enforcement, rather than interpersonally enforcing them. Future research should also investigate whether state punishments are less risky, difficult or awkward to implement than interpersonal ones and whether this varies across cultural tightness and looseness. Tight cultures tend to have more laws and more severe punishments of those laws, which may be reflected in everyday decision-making (Gelfand et al. 2011). Further, because the enactment of state punishment does not reflect the particular mental states of the enforcer, future research could investigate what reputational costs or benefits are afforded to those who seek out state (vs. interpersonal) punishment and those who enact it. Prior research suggests that punishments which are easy to physically implement and afford distance are preferred to those that are more involved (e.g., solitary confinement over cutting off an appendage; Bernhard, Cushman, and LeBaron 2022). This work suggests that state punishments should be easier to dole out, especially under social pressure, than other forms of punishment (see also Milgram 1974). Future research should test these claims and explore what state punishments communicate compared to other forms of punishment across different interpersonal and cultural contexts.

Grounding the study of punishment within the realm of state punishment opens up new areas of inquiry. People have expectations about how state rules and punishments should function, which are largely shared across cultures (Hannikainen et al. 2021). State rules may differ at the cognitive level from other kinds of rules in important ways. For example, people may be more likely to notice impossible to follow rules or laws compared to social norms, and given that moral values constrain what comes to mind as possible (Phillips and Cushman 2017), official rules may well do the same even for behaviours that are otherwise acceptable. Further, prior work has suggested that people often rely on agreement-based reasoning (i.e., contractualist reasoning) to decide when to follow rules (see Levine et al. 2024) and use loopholes to avoid outright disobedience and its associated costs (Qian et al. 2024), but it is possible that when it comes to state rules, the suite of mechanisms that underlie decisions to comply is unique. Future research should explore these themes and compare how expectations of rules and norms influence their differential enforcement.

2.2.2 | On Legitimacy and State Punishment

People also vary in their perceptions of legitimate institutions, and judgements of state punishments are likely influenced by levels of trust in the institutions that enforce the rules, and perceptions of their legitimacy (e.g., procedural fairness; Tyler 1997). Differential perceptions of legitimacy may in turn affect what is communicated by state punishment. Future research could also explore how individual differences in trust in the enforcing institutions affect judgements and decision-making the use of state punishment. Moreover, the enforcement of a longstanding rule may be perceived as particularly legitimate or longstanding rules may fail to catch up to societal change. It

is likely that the perceived legitimacy of the punishment that follows the violation of a longstanding rule is sensitive not only to whether the rule was literally broken (i.e., the letter of the law) but also the spirit or perceived intended purpose of the law and people's moral judgements of the proscribed behaviour (de Almeida, Struchiner, and Hannikainen 2023), which may change over time (Rozin 1999). Finally, future research could test when people see state punishment as legitimate, especially across different cultural contexts and historical timepoints where trust in governing institutions varies (see e.g., Atari, Henrich, and Schulz 2024). For example, rules vary in terms of how well they represent different people's views and values, and this may influence when people think state punishment, and perhaps most interestingly, state violence, is perceived to be legitimate across contexts.

3 | Conclusion

Here, we have argued that to understand the psychology of punishment, researchers must understand how people engage with state punishment as a special form of punishment, which makes up much of the punishment that people interact with in their lives. We have argued that state punishment is a special form of third-party punishment, pre-specified to occur after the violation of official rules and policies, carried out by people acting on behalf of a nation or government. Foregrounding state punishment in our psychology of punishment more generally highlights unique forms of violence, communication and the promotion of both cooperation and domination afforded by punishment.

Acknowledgements

We thank members of the PsyPhi Lab at the CUNY Graduate Center for helpful feedback on an earlier draft of this manuscript.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

We did not collect new data for this invited Outlook submission. As a result, we have neither an ethics statement to share about it nor links to the materials, code and codebook or any pre-registrations to share here or in the Methods section.

Endnotes

[1] These judgements are in keeping with Arendt's (1963) interpretation of the wrongdoing of Adolf Eichmann, as she famously highlighted the failure of the legal system to have a proper way to characterize those who carry out immoral orders.

[2] For instance, in the *Coates v. City of Cincinnati* (1971) case, the court deemed the city ordinance, which criminalized 'three or more persons to assemble' on a city sidewalk 'and there conduct themselves in a manner annoying to persons passing by' as unconstitutionally vague because the subjectivity of the term 'annoyance' made it 'unascertainable' (Mannheimer 2019, 1063). An average person cannot fully predict when, who and why someone may consider them annoying, making it an impossible rule to comply with.

[3] Despite the famous 'I know it when I see it' test for obscenity, the Miller test, established in *Miller v. California* (1973) outlines three key criteria for determining obscenity: (1) whether the average person, applying contemporary community standards, would find that the work appeals to prurient interest, (2) whether the work depicts or describes, in a patently offensive way, sexual conduct specifically defined by state law and (3) whether the work lacks serious literary, artistic, political or scientific value.

References

Andrews, K. 2009. "Understanding Norms Without a Theory of Mind." *Inquiry* 52, no. 5: 433–448. https://doi.org/10.1080/00201740903302584.

Arendt, H. 1963. *Eichmann in Jerusalem: A Report on the Banality of Evil*. London: Penguin Books.

Atari, M., J. Henrich, and J. Schulz. 2024. "Expanding the Remit of Psychology across Time and Space." Published Ahead of Print, May, 2024, https://doi.org/10.31234/osf.io/8atwz.

Baldassarri, D., and G. Grossman. 2011. "Centralized Sanctioning and Legitimate Authority Promote Cooperation in Humans." *Proceedings of the National Academy of Sciences* 108, no. 27: 11023–11027. https://doi.org/10.1073/pnas.1105456108.

Bernhard, R. M., F. A. Cushman, and H. LeBaron. 2022. "The Paradox of Aversive Punishment." PsyArXiv Preprints. Published Ahead of Print, August 6, 2024. https://doi.org/10.31234/osf.io/tcsve.

Bicchieri, C., E. Dimant, M. Gelfand, and S. Sonderegger. 2023. "Social Norms and Behavior Change: The Interdisciplinary Research Frontier." *Journal of Economic Behavior & Organization* 205: A4–A7. https://doi.org/10.1016/j.jebo.2022.11.007.

Boyd, R., and P. J. Richerson. 1992. "Punishment Allows the Evolution of Cooperation (or Anything Else) in Sizable Groups." *Ethology and Sociobiology* 13, no. 3: 171–195.

Buckholtz, J. W., C. L. Asplund, P. E. Dux, et al. 2008. "The Neural Correlates of Third-Party Punishment." *Neuron* 60, no. 5: 930–940. https://doi.org/10.1016/j.neuron.2008.10.016.

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.

Centola, D., J. Becker, D. Brackbill, and A. Baronchelli. 2018. "Experimental Evidence for Tipping Points in Social Convention." *Science* 360, no. 6393: 1116–1119.

Coates, v. City of Cincinnati, 402 U.S. 611 1971.

Cooperrider, K. 2022. "Animal Minds and Animal Morality." Broadcast. Retrieved September 15, 2024. https://podcasts.apple.com/gb/podcast/animal-minds-and-animal-morality/id1499167824?i=1000558848763.

Daston, L. 2022. *Rules: A Short History of What We Live by*. New Jersey: Princeton University Press.

de Almeida, G. D. F. C. F., N. Struchiner, and I. R. Hannikainen. 2023. "Rule Is a Dual Character Concept." *Cognition* 230: 105259. https://doi.org/10.1016/j.cognition.2022.105259.

Dennett, D. C. 1988/1993. "Quining Qualia." In *Readings in Philosophy and Cognitive Science*, edited by A. I. Goldman, 381–414. Cambridge: MIT Press. https://doi.org/10.7551/mitpress/5782.003.0022.

Fehr, E., and U. Fischbacher. 2004. "Third-Party Punishment and Social Norms." *Evolution and Human Behavior* 25, no. 2: 63–87. https://doi.org/10.1016/S1090-5138(04)00005-4.

Fehr, E., U. Fischbacher, and S. Gächter. 2002. "Strong Reciprocity, Human Cooperation, and the Enforcement of Social Norms." *Human Nature* 13, no. 1: 1–25. https://doi.org/10.1007/s12110-002-1012-7.

Fehr, E., and S. Gächter. 2000. "Cooperation and Punishment in Public Goods Experiments." *American Economic Review* 90, no. 4: 980–994. https://doi.org/10.1257/aer.90.4.980.

Fehr, E., and S. Gächter. 2002. "Altruistic Punishment in Humans." *Nature* 415, no. 6868: 137–140. https://doi.org/10.1038/415137a.

Foucault, M. 1977. *Discipline and Punish: The Birth of the Prison*. New York: Pantheon Books.

Fuller, L. L. 1964. *The Morality of Law*. New Haven: Yale University Press.

Gantman, A. P., A. Sternisko, P. M. Gollwitzer, G. Oettingen, and J. J. Van Bavel. 2020. "Allocating Moral Responsibility to Multiple Agents." *Journal of Experimental Social Psychology* 91: 104027. https://doi.org/10.1016/j.jesp.2020.104027.

Garcia, S. M., P. Chen, and M. T. Gordon. 2014. "The Letter Versus the Spirit of the Law: A Lay Perspective on Culpability." *Judgment and Decision Making* 9, no. 5: 479–490.

Gelfand, M. J., S. Gavrilets, and N. Nunn. 2024. "Norm Dynamics: Interdisciplinary Perspectives on Social Norm Emergence, Persistence, and Change." *Annual Review of Psychology* 75, no. 1: 341–378.

Gelfand, M. J., J. L. Raver, L. Nishii, et al. 2011. "Differences Between Tight and Loose Cultures: A 33-Nation Study." *Science* 332, no. 6033: 1100–1104.

Gintis, H., E. A. Smith, and S. Bowles. 2001. "Costly Signaling and Cooperation." *Journal of Theoretical Biology* 213, no. 1: 103–119. https://doi.org/10.1006/jtbi.2001.2406.

Graeber, D. 2015. *The Utopia of Rules: On Technology, Stupidity, and the Secret Joys of Bureaucracy*. Brooklyn: Melville House.

Griffin, T. 2021. "Comparing Expert Versus General Public Rationale for Death Penalty Support and Opposition: Is Expert Perspective on Capital Punishment Consistent With "Disciplined Retention"?." *Punishment & Society* 23, no. 4: 557–577. https://doi.org/10.1177/14624745211029370.

Hadfield, G. K., and B. R. Weingast. 2012. "What Is Law? A Coordination Model of the Characteristics of Legal Order." *Journal of Legal Analysis* 4, no. 2: 471–514.

Hampton, J. 1984. "The Moral Education Theory of Punishment." *Philosophy & Public Affairs* 13, no. 3: 208–238.

Hannikainen, I. R., K. P. Tobia, G. D. F. C. F. de Almeida, et al. 2021. "Are There Cross-Cultural Legal Principles? Modal Reasoning Uncovers Procedural Constraints on Law." *Cognitive Science* 45, no. 8: e13024. https://doi.org/10.1111/cogs.13024.

Hannikainen, I. R., K. P. Tobia, G. D. F. C. F. de Almeida, et al. 2022. "Coordination and Expertise Foster Legal Textualism." *Proceedings of the National Academy of Sciences* 119, no. 44: e2206531119. https://doi.org/10.1073/pnas.2206531119.

Henrich, J., and F. J. Gil-White. 2001. "The Evolution of Prestige: Freely Conferred Deference as a Mechanism for Enhancing the Benefits of Cultural Transmission." *Evolution and Human Behavior* 22, no. 3: 165–196. https://doi.org/10.1016/S1090-5138(00)00071-4.

Henrich, J., and M. Muthukrishna. 2021. "The Origins and Psychology of Human Cooperation." *Annual Review of Psychology* 72, no. 1: 207–240. https://doi.org/10.1146/annurev-psych-081920-042106.

Jara-Ettinger, J., and Y. Dunham. 2024. The Institutional Stance. https://doi.org/10.31234/osf.io/pefsx.

Jensen, K., J. Call, and M. Tomasello. 2007. "Chimpanzees Are Vengeful but Not Spiteful." *Proceedings of the National Academy of Sciences* 104, no. 32: 13046–13050. https://doi.org/10.1073/pnas.0705555104.

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.

Jordan, J. J., and N. S. Kteily. 2023. "How Reputation Does (and Does Not) Drive People to Punish Without Looking." *Proceedings of the National Academy of Sciences* 120, no. 28: e2302475120. https://doi.org/10.1073/pnas.2302475120.

Kahan, D. M. 1996. "What Do Alternative Sanctions Mean?" *University of Chicago Law Review* 63, no. 2: 591–653. https://doi.org/10.2307/1600237.

Keizer, K., S. Lindenberg, and L. Steg. 2008. "The Spreading of Disorder." *Science* 322, no. 5908: 1681–1685. https://doi.org/10.1126/science.1161405.

Knobe, J. 2010. "Person as Scientist, Person as Moralist." *Behavioral and Brain Sciences* 33, no. 4: 315–329. https://doi.org/10.1017/S0140525X10000907.

Legros, S., and B. Cislaghi. 2020. "Mapping the Social-Norms Literature: An Overview of Reviews." *Perspectives on Psychological Science* 15, no. 1: 62–80. https://doi.org/10.1177/1745691619866455.

Levine, S., M. Kleiman-Weiner, N. Chater, F. Cushman, and J. B. Tenenbaum. 2024. "When Rules Are Over-Ruled: Virtual Bargaining as a Contractualist Method of Moral Judgment." *Cognition* 250: 105790.

Lie-Panis, J., L. Fitouchi, N. Baumard, and J.-B. André. 2023. "The Social Leverage Effect: Institutions Transform Weak Reputation Effects Into Strong Incentives for Cooperation." *Evolution* 12, no. 51: e2408802121. https://doi.org/10.31234/osf.io/uftzb.

Locke, J. 1689/1988. *Two Treatises of Government*. Cambridge: Cambridge University Press. (Original work published 1689).

Malle, B. F., S. Guglielmo, and A. E. Monroe. 2014. "A Theory of Blame." *Psychological Inquiry* 25, no. 2: 147–186. https://doi.org/10.1080/1047840X.2014.877340.

Mannheimer, M. 2019. "Vagueness as Impossibility." *SSRN Electronic Journal* 28, no. 6: 1049–1114..

McAdams, R. H. 2015. *The Expressive Powers of Law: Theories and Limits*. Cambridge: Harvard University Press. https://doi.org/10.4159/harvard.9780674735965.

Milgram, S. 1974. *Obedience to Authority: An Experimental View*. New York: Harper & Row.

Miller, v. California, 413 U.S. 15 1973.

Mills, C. W. 2005. ""Ideal Theory" as Ideology." *Hypatia* 20, no. 3: 165–183. https://doi.org/10.1111/j.1527-2001.2005.tb00493.x.

Muthukrishna, M., P. Francois, S. Pourahmadi, and J. Henrich. 2017. "Corrupting Cooperation and How Anti-Corruption Strategies May Backfire." *Nature Human Behaviour* 1, no. 7: 0138. https://doi.org/10.1038/s41562-017-0138.

Ostrom, E. 1990. *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge: Cambridge University Press.

Ozono, H., N. Jin, M. Watabe, and K. Shimizu. 2016. "Solving the Second-Order Free Rider Problem in a Public Goods Game: An Experiment Using a Leader Support System." *Scientific Reports* 6, no. 1: 38349. https://doi.org/10.1038/srep38349.

Pew Research Center. 2021. "Most Americans Favor the Death Penalty Despite Concerns About Its Administration." https://www.pewresearch.org/politics/2021/06/02/most-americans-favor-the-death-penalty-despite-concerns-about-its-administration/.

Phillips, J., and F. Cushman. 2017. "Morality Constrains the Default Representation of What Is Possible." *Proceedings of the National Academy of Sciences* 114, no. 18: 4649–4654.

Qian, P., S. Bridgers, M. Taliaferro, K. Parece, and T. D. Ullman. 2024. "Ambivalence by Design: A Computational Account of Loopholes." *Cognition* 252: 105914. https://doi.org/10.1016/j.cognition.2024.105914.

Raihani, N. J., and R. Bshary. 2019. "Punishment: One Tool, Many Uses." *Evolutionary Human Sciences* 1: e12. https://doi.org/10.1017/ehs.2019.12.

Riedl, K., K. Jensen, J. Call, and M. Tomasello. 2012. "No Third-Party Punishment in Chimpanzees." *Proceedings of the National Academy of Sciences* 109, no. 37: 14824–14829. https://doi.org/10.1073/pnas.1203179109.

Rozin, P. 1999. "The Process of Moralization." *Psychological Science* 10, no. 3: 218–221.

Sarin, A., M. K. Ho, J. W. Martin, and F. A. Cushman. 2021. "Punishment Is Organized Around Principles of Communicative Inference." *Cognition* 208: 104544. https://doi.org/10.1016/j.cognition.2020.104544.

Saxe, R., and N. Kanwisher. 2013. "People Thinking About Thinking People: The Role of the Temporo-Parietal Junction in "Theory of Mind"." In *Social neuroscience*, 171–182. Psychology Press.

Scott, J. C. 2020. *Seeing Like a State*. New Haven: Yale University Press.

Struchiner, N., I. R. Hannikainen, and G. D. F. C. F. de Almeida. 2020. "An Experimental Guide to Vehicles in the Park." *Judgment and Decision Making* 15, no. 3: 312–329.

Sunshine, J., and T. R. Tyler. 2003. "The Role of Procedural Justice and Legitimacy in Shaping Public Support for Policing." *Law & Society Review* 37, no. 3: 513–548.

Sunstein, C. R. 1996. "On the Expressive Function of Law." *University of Pennsylvania Law Review* 144, no. 5: 2021–2053. https://doi.org/10.2307/3312647.

Transportation Security Administration. 2022. Enforcement Sanction Guidance Policy. Washington, DC: United States Department of Homeland Security (DHS).

Turri, J. 2019. "Excuse Validation: A Cross-Cultural Study." *Cognitive Science* 43, no. 8: e12748.

Tyler, T. R. 1997. "Procedural Fairness and Compliance With the Law." *Swiss Journal of Economics and Statistics* 133, no. II: 219–240.

Tyler, T. R. 2003. "Procedural Justice, Legitimacy, and the Effective Rule of Law." *Crime and Justice* 30: 283–357. https://doi.org/10.1086/652233.

Weber, M. 1919. "Politics as a Vocation." In *Max Weber: Selections in Translation*, edited by W. G. Runciman, translated by E. Matthews, 212–225. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511810831.

Wheeler, N. E., S. Allidina, E. U. Long, S. P. Schneider, I. J. Haas, and W. A. Cunningham. 2020. "Ideology and Predictive Processing: Coordination, Bias, and Polarization in Socially Constrained Error Minimization." *Current Opinion in Behavioral Sciences* 34: 192–198. https://doi.org/10.1016/j.cobeha.2020.05.002.

Wylie, J., and A. Gantman. 2023. "Doesn't Everybody Jaywalk? On Codified Rules That Are Seldom Followed and Selectively Punished." *Cognition* 231: 105323. https://doi.org/10.1016/j.cognition.2022.105323.

Wylie, J., and A. Gantman. 2024. "Cooperation, Domination: Twin Functions of Third-Party Punishment." *Social and Personality Psychology Compass* 18, no. 8: e12992. https://doi.org/10.1111/spc3.12992.

Wylie, J., K. L. Milless, J. Sciarappo, and A. Gantman. 2024. The Biased Enforcement of Rarely Followed Rules. Personality and Social Psychology Bulletin. https://doi.org/10.1177/01461672241252853.