

- Xu, Z., Zhu, R., Zhang, S., Zhang, S., Liang, Z., Mai, X., & Liu, C. (2022). Mortality salience enhances neural activities related to guilt and shame when recalling the past. *Cerebral Cortex*, 32(22), 5145–5162. <https://doi.org/10.1093/cercor/bhac004>
- Yu, H., Hu, J., Hu, L., & Zhou, X. (2014). The voice of conscience: Neural bases of interpersonal guilt and compensation. *Social Cognitive and Affective Neuroscience*, 9(8), 1150–1158. <https://doi.org/10.1093/scan/nst090>
- Zhu, R., Feng, C., Zhang, S., Mai, X., & Liu, C. (2019). Differentiating guilt and shame in an interpersonal context with univariate activation and multivariate pattern analyses. *NeuroImage*, 186, 476–486. <https://doi.org/10.1016/j.neuroimage.2018.11.012>
- Zhu, R., Jin, T., Shen, X., Zhang, S., Mai, X., & Liu, C. (2017). Relational utility affects self-punishment in direct and indirect reciprocity situations. *Social Psychology*, 48(1), 19–27. <https://doi.org/10.1027/1864-9335/a000291>
- Zhu, R., Shen, X., Tang, H., Ye, P., Wang, H., Mai, X., & Liu, C. (2017). Self-punishment promotes forgiveness in the direct and indirect reciprocity contexts. *Psychological Reports*, 120(3), 408–422. <https://doi.org/10.1177/0033294117697087>

Authors' Response

The puritanical moral contract: Purity, cooperation, and the architecture of the moral mind

Léo Fitouchi , Jean-Baptiste André*
and Nicolas Baumard*

Département d'études cognitives, Institut Jean Nicod, École normale supérieure, Université PSL, EHESS, CNRS, Paris, France
leo.fitouchi@gmail.com; <https://sites.google.com/view/leofitouchi/home>
jeanbatpisteandre@gmail.com; <http://jb.homepage.free.fr/>
nbaumard@gmail.com; <https://nicolasbaumard.org/>

doi:10.1017/S0140525X23001188, e322

Abstract

Commentators raise fundamental questions about the notion of purity (sect. R1), the architecture of moral cognition (sect. R2), the functional relationship between morality and cooperation (sect. R3), the role of folk-theories of self-control in moral judgment (sect. R4), and the cultural variation of morality (sect. R5). In our response, we address all these issues by clarifying our theory of puritanism, responding to counter-arguments, and incorporating welcome corrections and extensions.

We are immensely grateful to all commentators for their interest, thought-provoking arguments, and the fascinating discussion they open up on the nature of morality. We are thrilled that most theories of morality are represented in the commentaries, including moral foundations theory (Graham, Atari, Dehghani, & Haidt [Graham et al.]), dyadic morality (DiMaggio, Gray, & Kachanoff [DiMaggio et al.]), morality as cooperation (Curry & Sznycer), as well as the side-taking hypothesis and related accounts (DeScioli & Kurzban; Moon; Tybur & Lieberman).

The purity controversy has structured moral psychology for decades, and for good reason (Gray, DiMaggio, Schein, & Kachanoff, 2022). Purity is at the junction of two heated debates between the aforementioned theories of morality:

- (1) *The monism–pluralism debate*. Are all moral judgments, despite the diversity of their content (e.g., purity, fairness, authority), produced by a single computational device? Yes, according to monist theories – and purity is no exception (DiMaggio et al.; DeScioli & Kurzban, 2009; target article). Pluralists disagree: Purity, in their view, reveals the inability of monist models to explain the diversity and complexity of moral judgments (Curry & Sznycer; Graham et al.; Weinstein & Baldwin).
- (2) *The cooperation debate*. Did moral cognition evolve exclusively for cooperation? We and others claim so (Curry & Sznycer; Kurdoglu; Murray, Amaya, & Jiménez-Leal [Murray et al.]). But many disagree: Purity moralizations, they argue, reveal that other adaptive challenges, such as pathogen avoidance (Graham et al.), coordination for side-taking in disputes (DeScioli & Kurzban), and self-serving use of moral principles (Moon; Tybur & Lieberman), have shaped the moral mind in our evolutionary history.

In this context, our target article aimed to show that moralizations of purity, often taken as a critical argument against both monist (Graham et al., 2013) and cooperation-based theories (Smith & Kurzban, 2019), pose a problem for neither types of theories.

Expectedly, both claims proved controversial. Pluralists think we are too reductionist. Positing more moral cognitive systems, they argue, allows better explaining morality in general and puritanism in particular. Opponents of cooperation-centric views, meanwhile, think we're too naive. Puritanism is not about cooperation, they argue, but about oppressive coercions, manipulative condemnations, and cruel punishments. As if sorting out these issues weren't difficult enough, the task is further complicated by the general confusion about what we're supposed to explain when we talk about “purity” (DiMaggio et al.; Kollareth & Russell).

We thus begin by clarifying our explanatory target – puritanical morality – by distinguishing it from other purity-related moralizations (sect. R1). This sets the stage for addressing the monism–pluralism debate (sect. R2) and the cooperation debate (sect. R3). We finally discuss the role of folk-theories of self-control in puritanical moral judgments (sect. R4), as well as cultural variations in puritanical values (sect. R5).

R1. Puritanism and purity: Clarifying explanatory targets

In evaluating our model, many commentators discussed purity violations such as incest (Tybur & Lieberman), atheism, blasphemy (DeScioli & Kurzban), food taboos (DiMaggio et al.; Giner-Sorolla & Myers), premarital sex (Weinstein & Baldwin), homosexuality (Giner-Sorolla & Myers; Tybur & Lieberman; DeScioli & Kurzban), rolling in urine (DiMaggio et al.), eating the family's dead pet dog (Murray et al.), or defecating on someone's grave (Murray et al.).

Most of these behaviors, however, were not clearly included in our definition of puritanical morality (target article, sect. 1). This confusion is natural given that purity is a fuzzy concept and that puritanism is a subset of purity. Before turning to more substantial debates (sects. R2 and R3), let us try to bring some order to this conceptual mess. We do so by distinguishing puritanical morality from purity (sect. R1.2), sexual morality (sect. R1.3), and the morality of the historical Puritans (sect. R1.4).

*Co-last authors.

R1.1. Can we (please) stop talking about “purity?”

As **Graham et al.** note, we think purity has brought a lot to moral psychology, by drawing attention to cultural variation and moral intuitionism. We introduced the notion of puritanical morality, however, because purity seemed too vague of a notion to provide a good explanatory target. Our target article converges with recent recommendations to abandon or at least reconceptualize the notion of purity in moral psychology (Crone, 2022; Gray et al., 2022; Kollareth, Brownell, Duran, & Russell, 2022). We and other commentators see three reasons to do so.

First, purity is poorly defined. In their recent review, **DiMaggio et al.** show that purity is understood in nine different ways (Gray et al., 2022). Rather than being given a proper analytic definition, purity functions as an intuitive label for violations that loosely relate to sex or food or religion or pathogens. As a result, “nobody even knows what exactly purity is” (DiMaggio et al., para. 1). If we don’t know what we’re supposed to explain, we’re unlikely to advance our causal understanding and may be condemned to sterile debates.

Second, purity is not a distinct cognitive domain (**Kollareth & Russell**). Despite its fuzzy contours, one reason to keep the notion of purity may be that all purity violations trigger a common and distinctive cognitive response. Kollareth & Russell, however, review compelling evidence that three popular criteria for carving the purity domain fail to distinguish purity from other violations. Purity is (1) not tied to a specific moral emotion; (2) is not perceived as “tainting the soul” more than other violations; and (3) is not less sensitive to the actor’s intentions than other moral judgments (Kollareth & Russell).

Third, many purity scenarios are so weird that they distract us from real-world moral concerns (target article, sect. 2.1). Think about eating pizza off a corpse (Clifford, Iyengar, Cabeza, & Sinnott-Armstrong, 2015) or touching poop barehanded (Dungan, Chakroff, & Young, 2017). People don’t cite these behaviors as typical moral violations (Gray & Keeney, 2015). And it’s not even clear whether people perceive them as really immoral (Kollareth et al., 2022).

Given these problems, it may be better to focus our explanatory efforts on (1) more restricted clusters of moral judgments, that (2) better carve moral culture at its cognitive joints, and (3) better reflect real-world moral concerns. As **DiMaggio et al.** note, our target article aimed to “tackle [this] mess of purity” by attempting to delineate one such subset of purity concerns that is (hopefully) more workable.

R1.2. Puritanical morality is not purity

By puritanism, we referred to a subset of purity that seemed to share a common feature: the moralization of asceticism. This cluster of concerns seemed to make sense in light of human psychology. Sex, food, drugs, alcohol, laziness, and ecstatic music and dances, all give us intense pleasures – the exhilarating shots of dopamine you get when you orgasm or eat sugar. Puritanism, at its very core, is a moral fear of these hedonic states. Pleasure, in puritanical eyes, is a gateway to excess, addiction, uncontrollable cravings, and irresponsible self-gratification. If you taste it, you’ll seek for more, even when it means neglecting obligations – it’s just too good! If you want to escape the hedonic trap and become a better person, you must learn to tame the flesh and resist worldly temptations. Hence the prescription of ascetic moderation.

Many canonical purity scenarios don’t clearly instantiate these ascetic concerns. We agree with commentators that sibling incest (**Tybur & Lieberman**), eating the family’s dead dog (**Murray et al.**), defecating on someone’s grave (Murray et al.), disgusting behaviors such as rolling in urine (**DiMaggio et al.**), and food taboos after the death of a loved one among Hindu Brahmins (DiMaggio et al.) are most often not moralized because perceived as endangering self-control – although this may be worth actually testing.

We share **DiMaggio et al.**’s view that these judgments stem from perceptions of (unfair) harm not mediated by beliefs about self-control (Fitouchi, André, & Baumard, *in press*). As DiMaggio et al. note, not respecting a special diet after the death of a loved one, although seemingly harmless to Western researchers, was perceived by Orissa Hindu Brahmins as harming the soul of the deceased by delaying their reincarnation (Shweder, 2012). Similarly, people likely perceive defecating on someone’s grave (**Murray et al.**) as offending the deceased or their family, and probably calculate that everyone is worse off in a society where everyone shits on others’ graves compared to a society where nobody does (see sect. R2.1). Julie and Mark’s sibling incest (**Tybur & Lieberman**), meanwhile, fails to convince participants that the action they’re judging is really harmless, and these perceptions of harm predict their condemnation of the act better than disgust (Royzman, Kim, & Leeman, 2015). We like, however, **Kurdoglu**’s suggestion that concerns for self-control might nonetheless underlie the moralization of incest, and would very much like to see this idea tested.

R1.3. Puritanical morality is not sexual morality

Many commentaries assess the explanatory power of our model by discussing sexual morality in general (**Royzman & Borislow; Weinstein & Baldwin; Szocik**). Puritanism, however, is not sexual morality (Fig. R1). Sexual puritanism captures a very specific type of sexual morality, which condemns, not particular sex acts such as adultery (**Syme**), premarital sex (Weinstein & Baldwin; **Kurdoglu**), or restrictions on female sexuality (**Barenthin; Royzman & Borislow; Szocik**), but the very fact of *taking pleasure* in sex (Dabhoiwala, 2012; Greenberg & Bystry, 1982; Le Goff, 1984; Suiming, 1998). This is why we exemplified sexual puritanism by the moralization of masturbation and the prescription, even in marriage, that sex should always be consumed in moderation, never in a sensual way, and always for the necessity of procreation rather than to enjoy its pleasures (Dabhoiwala, 2012; Seidman, 1990; Fig. R1).

Sexual morality includes many moral judgments unrelated to this ascetic restriction of sex simply because it’s pleasurable (Fig. R1). In particular, we did not argue that adultery and premarital sex are condemned because they are perceived as endangering self-control (**Weinstein & Baldwin**). Adultery imposes direct costs on the cheated partner (target article, sect. 1.2, endnote 1). Premarital sex, in many societies, imposes direct costs on families by leading to unwanted marriages, costly pregnancies out of wedlock, or decreasing a daughter’s value on the matrimonial market (target article, sect. 3.2, endnote 3). As **Szocik** and **Barenthin** note, many restrictions on women’s chastity and fidelity arise from men’s interest to control women’s sexuality (see sect. R3.1). This often leads to a moral contract between men only, in which they promise each other not to covet each other’s wives (Dabhoiwala, 2012; see also Szocik).

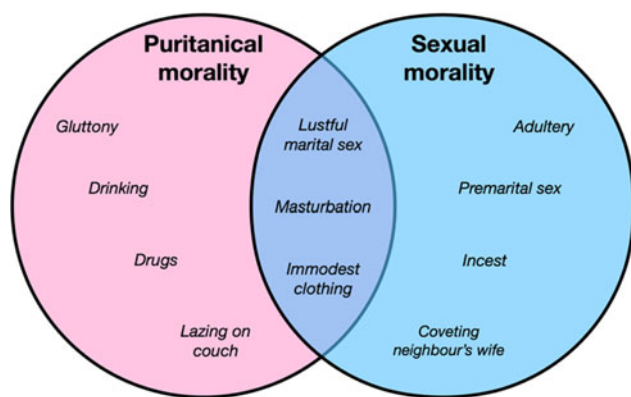


Figure R1. Clarifying the relationship between puritanical morality and sexual morality.

Our theory does not attempt to explain moralizations of these sexual behaviors that are intrinsically harmful. In our model, adultery, premarital sex, and sex with someone else's spouse are precisely among the *harmful* temptations to which puritanism seeks to improve your resistance, by training you to resist even *harmless* sexual pleasures, such as masturbation and lustful marital sexuality¹ (Fig. R2). Note that other bodily pleasures, too, can be both directly harmful and intrinsically harmless. Eating too much food *from the common pot* directly harms other people by depriving them from resources they deserve. Indulging in laziness or intoxicants *when you're supposed to do your part of collective work* amounts to directly free-ride on others' contributions – as Syme rightly notes. Again, these are *directly harmful* temptations to which puritanism seeks to improve your resistance, by limiting even *harmless* indulgences, such as eagerly eating *your own food*, and lazing on the couch *when you don't have to work* (Fig. R2).

R1.4. Puritanical morality is not the morality of the historical Puritans

DeScioli & Kurzban argue that our model fails to explain “puritanism” because behaviors condemned by the historical Puritans, such as atheism, blasphemy, witchcraft, or carrying wood on Sunday, seem unrelated to self-control. Again, these behaviors don't clearly instantiate the ascetic concerns at the heart of puritanical morality as we define it. We did not define puritanical morality as “the set of moral values held by the historical Puritans” (target article, sect. 1). The Puritans moralized many things: Not only bodily pleasures and ascetic restraint, but also non-puritanical concerns such as theft, treason, murder, justice, charity, humility, peacefulness, and many other values (Hall, 2012; Merrill, 1945). Conversely, puritanical concerns are found, not only among the historical Puritans, but across world religions more generally (target article, sect. 1). It doesn't seem helpful, then, to define puritanical morality as the morality of the historical Puritans.

With these clarifications in mind, we can turn to more substantial debates about the architecture of the moral mind (sect. R2), and the adaptive function of moral cognition (sect. R3).

R2. Puritanism and the moral mind: One or many moral modules?

Our target article argued that all moral judgments – including puritanical ones – are produced by a single, functionally unified cognitive system. Several commentaries call for a more pluralistic approach to moral cognition, arguing that puritanism cannot be reduced to “harm or fairness” (Curry & Sznycer; Graham et al.; Weinstein & Bladwin; see also Buchtel). In this section, we defend moral monism by clarifying our view of the computational logic of moral cognition (sect. R2.1), and by using this logic

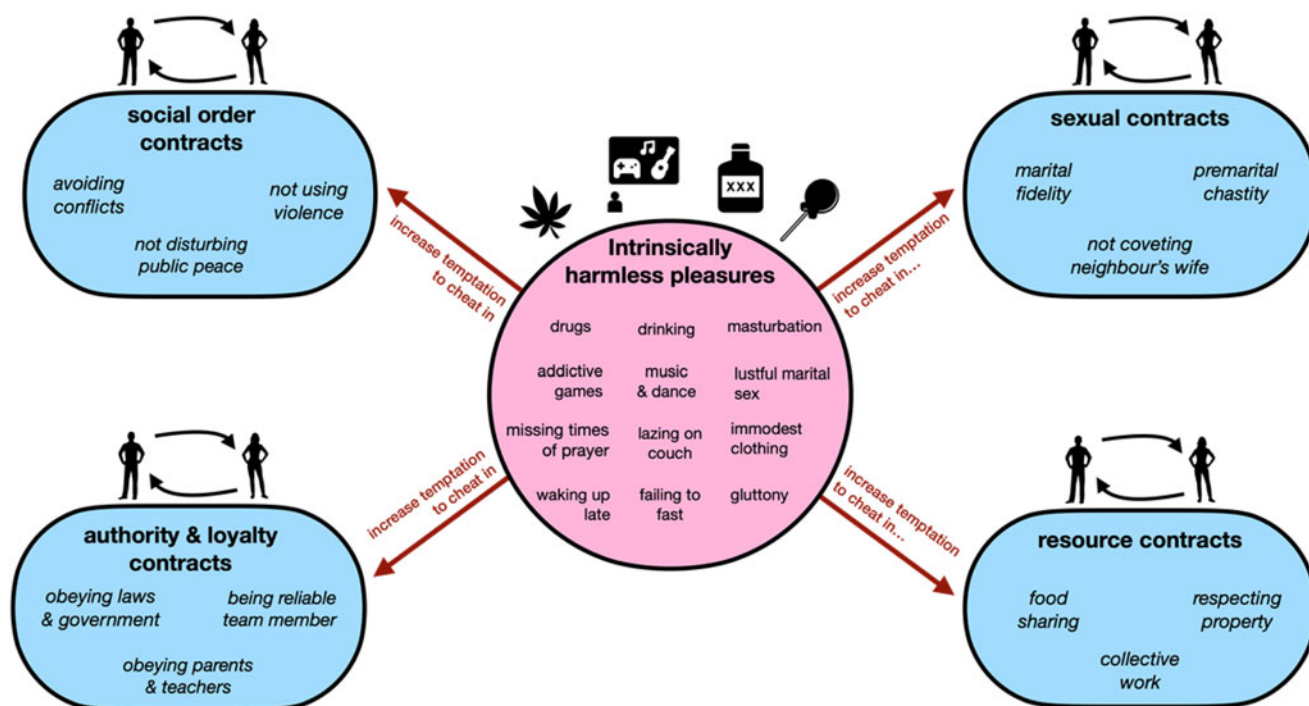


Figure R2. Perceived relationships between harmless pleasures and directly harmful behaviors in reciprocal contracts such as marital fidelity, food sharing, and social order.

to clarify the mechanisms of puritanical moral judgment (sect. R2.2) and moral emotions (sect. R2.3).

R2.1. A single computational device – calculating reciprocal contracts – explains moral judgments across domains

Our article started from the vague idea that “moral cognition evolved for cooperation.” We did so because many theories of morality agree with different variants of this claim (Boehm, 2012; Curry, 2016; Haidt, 2012; Tomasello, 2020). Yet our account of puritanism builds on one particular cooperation-based theory – the evolutionary contractualist theory of morality (André, Fitouchi, Debove, & Baumard, 2022; for earlier versions, see Baumard, André, & Sperber, 2013). According to this account, the computation moral cognition evolved to perform is strictly the same across domains of social interaction. It amounts to calculate reciprocal obligations that specify what each partner ought to do – despite having a temptation to cheat – to maximize the mutual benefits of their interaction (André et al., 2022; Fitouchi et al., in press). Before turning to the particular case of puritanism, it is necessary to clarify why we disagree with more pluralistic theories of morality.

Graham et al. make a detailed case for moral pluralism. The fact that moral cognition functions for cooperation, they argue, does not imply that the mind contains only one moral calculator to realize this function. Humans cooperate for many different purposes – from parenting to coalitions to resource production. Thus, moral cognition contains several domain-specific calculators, each tailored to one of these specific cooperation problems. These calculators include reciprocity, but also status hierarchies, coalitional psychology, and the behavioral immune system (Fig. R3). Reducing morality to harm or fairness, in their view, not only fails to explain the full breadth of morality in general (e.g., loyalty, authority), but also fails to fully explain puritanism in particular, because moralizations of bodily pleasures are associated with concerns for loyalty and authority (Goenka & Thomas, 2022; Mooijman et al., 2018).

We fully agree that puritanism aims to promote, not only fair distributions of resources, but also loyalty to coalition partners and obedience to authorities, as well as many other cooperative behaviors (target article, sect. 3). But the error of pluralistic theories is precisely to equate the plurality of moral concerns (e.g., fairness, loyalty, authority) with a plurality of moral cognitive systems (e.g., reciprocity, coalitional psychology, status hierarchies; Fig. R3). Where pluralistic models confine reciprocity to a tiny part of the moral mind, we argue that moralizations of loyalty, authority, and any other behavior, arise from the same computations of reciprocal obligations that produce fairness concerns (André et al., 2022; Fig. R3). More than that, we claim that judgments of loyalty, authority, and purity, are poorly explained by the non-reciprocal systems that pluralists have added into the moral mind to explain them. How can we tell? Look at the precise logic of moral intuitions in all these domains.

Take authority. Pluralists argue that moralizations of obedience to authority arise from systems evolved to navigate status hierarchies, akin to those evolved in nonhuman species, and distinct from reciprocity (Graham et al.; Curry & Sznycer). As Curry & Sznycer explicate, nonhuman hierarchies emerge from asymmetric hawk-dove interactions, where a weaker individual submits to a dominant based on cues of the dominant’s likelihood to win fights over contested resources. This allows both the subordinate and the dominant to avoid the mutual costs of conflict, as the subordinate leaves the resource to the dominant.

Yet this idea makes aberrant predictions. If moral intuitions about authority were produced by a calculator dedicated to this adaptive problem, the simple fact of being more likely to win a contest – like a dominant gorilla in a primate hierarchy – should give people a moral right to others’ obedience. Brute force, in other words, should be the only source of legitimate power. This completely contradicts people’s moral intuitions about authority. The core feature of authority as a moral relationship is precisely its difference from coercion (Saxe, 2022; Tyler, 2006). The ability to win conflicts over contested resources does not make a chief, a boss, or a teacher deserve his followers’

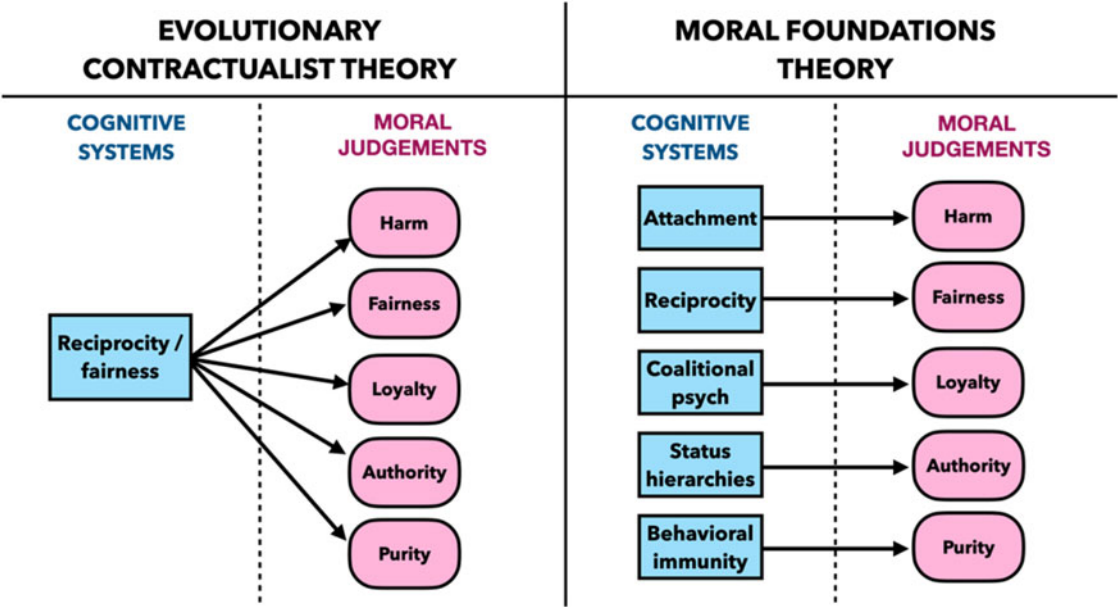


Figure R3. Distinction between the evolutionary contractualist theory of morality (André et al., 2022) and moral foundations theory (Graham et al., 2018).

obedience. It does allow him to *force* them to fulfill his desires. But that is precisely judged as an *abuse of power*, not as a moral right. Of course, subordinates will submit, superficially appearing to “respect authority,” but they will do so out of fear of the whip, not out of a moral obligation to obey.

Rather than from hawk–dove interactions, moral intuitions about authority emerge from the fact that authority is a *reciprocal contract* as any other (Price & Van Vugt, 2015). Leaders provide, at a cost to themselves, benefits to followers by working out complicated decisions, mediating disputes, or coordinating collective action (Glowacki & von Rueden, 2015; Hagen & Garfield, 2019; Price & Van Vugt, 2015). In return, followers provide leaders with status, resources, and decision-making power (Glowacki & von Rueden, 2015; Price & Van Vugt, 2015). Each follower accepts to give up some of their freedom to follow orders so that everyone can benefit from better decisions and more efficient collective action.

As in any reciprocal contract, what people consider morally wrong is to *cheat*. The leader can abuse his power to unfairly advance his interests at the expense of followers. This amounts to cheating: Taking the benefits of followers’ cooperation (obedience) while failing to fulfill his own part of the contract (making decisions that benefit everyone) (Tooby, Cosmides, & Price, 2006). Another key marker of reciprocity is the *conditionality* of people’s obligation to cooperate. If the leader neglects the interests of his people, people don’t feel morally obliged toward him in return (Tyler, 2006). They will submit if forced to. But they’ll nonetheless consider the tyrant morally corrupt. And they’ll choose another leader as soon as they can (Van Vugt, Jepson, Hart, & De Cremer, 2004), just as they seek better cooperation partners when cheated in other relationships (Baumard et al., 2013). In other words, authorities must be... fair. Wait, wasn’t fairness a distinct foundation?

The same holds for loyalty, which Graham et al. and DeScioli & Kurzban (para. 4) distinguish from reciprocal cooperation. Coalitions are nothing but *n*-person reciprocal exchanges (Tooby et al., 2006). In zero-sum competition between groups, helping a rival coalition means harming your partners (Boyer, Firat, & van Leeuwen, 2015). So coalition partners can maximize mutual benefit if each refrains from helping rival groups or shifting alliances. Yet each partner has a temptation to cheat. I could make money, for example, by selling strategic information to the group we’re fighting. If all my partners did that, however, we would lose the war and all be worse off. Here again, the immoral behavior – betrayal – amounts to cheating in a reciprocal contract: Taking the benefits of others’ refusal to trade with outgroups while not myself paying this cost. And loyalty intuitions obey the conditionality of reciprocal obligations: If all your partners shamelessly betrayed you, do you still owe them to be loyal?

To be clear, we do not deny that people have a coalitional psychology (Graham et al.), in the sense of domain-specific adaptations for *detecting* alliances or *recruiting* coalitional support (Pietraszewski, Curry, Petersen, Cosmides, & Tooby, 2015). But this is not the same thing as calculating “what we owe to each other” (Scanlon, 2000) within a coalition. The requirements of this properly moral computation are the same whether our collective action is about sharing food or competing with rivals or building a house together (André et al., 2022).

To morally judge a behavior, the moral calculator takes as inputs the *costs* and *benefits* that the behavior implies for each partner, and computes whether it would be mutually advantageous if all partners behaved in this way (André et al., 2022). What matters for this calculation is the amount of costs each

partner pays and the benefits they receive, regardless of the specific behaviors that generate these costs and benefits. If each of us shares food when others are hungry, we each end up with more benefits than costs. That’s because the cost of sharing when you have plenty is smaller than the benefit of being helped in return when you’re hungry. So we *owe* to each other to share food, and stinginess is *wrong* (food-sharing contract). In other situations, if each of us obeys orders, we each end up with more benefits than costs. That’s because the cost of obeying orders when we don’t like them is lower than the benefit we can get, for example, from the more efficient organization allowed by everyone’s obedience (obedience contract). The list of moral contracts could go on and on. If each of us refrains from using violence to his advantage, we’ll all benefit from the myriad of activities enabled by peace (social order contract). There is not one, or two, or five, or six moral foundations, but as many moral concerns as there are mutually beneficial contracts, which depend on the infinite, parametrical variation of costs and benefits implied by each behavior in every situation. Morality is not a set of foundations, but one flexible and open-ended calculation based on varying inputs (André et al., 2022).

With that in mind, we can turn to the particular contract that was the focus of our target article.

R2.2. The puritanical moral contract

R2.2.1. Fairness, not harm

The same computations of mutual benefit, we argue, explain puritanical morality. In our target article, we wrote that bodily pleasures are moralized when perceived to facilitate “harm” or “antisocial behaviors.” As Graham et al. rightly note, that was too vague: Is puritanism produced by perceptions of harm or fairness computations? The short answer is fairness.

We largely agree with DiMaggio et al.’s theory of dyadic morality that purity violations are condemned because they are perceived to indirectly cause harm. In our view, however, the immorality of an action does not depend on whether an agent *harms* a patient *per se*, but on whether an agent *cheats* a patient (André et al., 2022). Because cheating implies harming someone, harm and cheating often co-occur, giving the impression that morality is fundamentally about harm. But not all harm implies cheating. Sometimes, harming others is simply regrettable but not morally wrong.

Take breaking up with someone you’ve been dating for a few weeks and who happens to have feelings for you (Royzman & Borislow, 2022). Despite being harmful to your date – you broke their heart – people don’t find your behavior immoral (Royzman & Borislow, 2022). This is because it would not be mutually beneficial to expect everyone not to break up in such situations. Each of us would have to remain stuck forever with whomever we just started dating (huge cost). Moral cognition calculates that we are all better off allowing each other to break up sometimes (huge benefit) even if it means paying the smaller cost of occasional heartbreak (when others leave us). Thus, breaking up in this context is harmful but does not amount to cheating someone in a reciprocal contract. Accordingly, moral cognition does not calculate that this behavior is immoral (Royzman & Borislow, 2022).

The same holds, we argue, for puritanical moral judgments. Bodily pleasures are perceived to facilitate harmful behaviors, but are not moralized because of this harm in itself. The relevant question is not whether costs are imposed *per se*, but whether the net effect of costs and benefits received by the perpetrator would be positive if everyone behaved like the perpetrator – consistent

with DiMaggio et al.'s remarks about universalization (Levine, Kleiman-Weiner, Schulz, Tenenbaum, & Cushman, 2020). In other words, would we all be better off if we allowed each other the pleasure of drinking, but also suffered the costs of others' drinking (e.g., violence, adultery, lazy free-riding)? Or would we all be better off if we each paid the costs of abstinence yet ensured the benefits of safe streets and productive work (because others don't drink)? When people believe that unrestrained indulgence would result in huge costs, we argue, the moral computation outputs the puritanical moral contract, which goes something like this:

The puritanical moral contract: We owe it to each other to bear the costs of ascetic habits and rigorous training in self-control, because if everyone makes this effort, we'll all secure the greater benefits of an orderly and peaceful society. In this contract, gluttons, drunkards, and lustful sex-addicts are *cheaters*. They unfairly take the benefits of others' restraint without paying the costs of mutually beneficial discipline.

R2.2.2. A broader theory of cooperation doesn't help explain puritanism

Puritanical morality, Curry & Sznycer argue, is better explained by considering more cooperative problems than reciprocal contracts. They point to resolution of conflicts by ritual contests in hawk-dove interactions. When competing over a resource, contestants display cues of their likelihood to win a fight were it to escalate. This allows them to avoid the mutual costs of conflict, as the weaker individual withdraws from the contest (Smith & Parker, 1976). Because one area of human conflict is the competition over mates, Curry & Sznycer argue that traits signaling people's mate-value will be moralized because they help resolve conflicts over mates, by signaling contestants' relative ability to win the mating competition.

From this, however, Curry & Sznycer jump to an apparently unwarranted prediction, namely that cues of *low* mate-value, such as infidelity and masturbation, should be considered morally *bad*. This prediction does not follow from the hawk-dove interaction they describe. In a contest over mates, displaying *low* mate-value means playing dove – that is, signaling that I'm *unlikely* to win the mating competition. Yet playing dove is precisely a cooperative behavior: It allows preventing the conflict by leaving the resource to the hawk – people more likely to win the conflict. It is precisely when nobody displays inferiority that the conflict escalates. Morality as cooperation, it seems to us, thus predicts the opposite of what Curry & Sznycer make it say. Because cues of low mate-value allow resolving the conflict, morality as cooperation predicts that they should be considered morally *good* – not morally *bad*.

Curry & Sznycer present data that contradict this prediction. They find that perceiving purity violations, such as masturbation, as sexually unattractive correlates with considering them morally *bad*. This association is interesting, but in our view, it contradicts Curry & Sznycer's conflict-resolution theory when its predictions are carefully derived. Consistent with moral disciplining, however, Curry & Sznycer find that perceiving purity violations as indicating lack of self-control predicts their moral condemnation.

R2.3. Puritanism and moral emotions

R2.3.1. Disgust does not generate puritanical moral judgments

Our target article reviewed evidence against a role of disgust in puritanical condemnations (sect. 2.1). Graham et al., however,

argue that the writings of the puritans suggest that they “use disgust (as when contemplating a corpse) as a means of making moral progress by breaking one's attachments to one's own body.” We agree! But *using disgust* to make moral progress is not the same thing as *generating a moral judgment based on disgust*.

A psychological theory of morality is supposed to identify the cognitive mechanisms that *generate* moral judgments. Generating a moral judgment means taking as input a *non-moral* representation (e.g., cues of pathogens, costs imposed on other people), and producing, based on computations of this non-moral material, an output representation that does have moral content, such as “masturbation is *wrong*” (Fitouchi et al., *in press*). The current state of the evidence indicates that disgust is not such a mechanism: Simply perceiving a behavior as disgusting does not in itself generate the representation that it is wrong (Kollareth & Russell; Fitouchi et al., *in press*; Piazza, Landy, Chakroff, Young, & Wasserman, 2018).

Of course, once mechanisms other than disgust have produced the moral representation that masturbation is wrong, people can use disgust to nudge themselves into avoiding to behave immorally. But this does mean that it was disgust that produced their moral judgment. People can also self-inflict pain after masturbating to avoid doing it again. This does not imply that pain is a moral calculator – and the same holds for disgust.

Tybur & Lieberman argue that disgust may serve as an input to moral calculators rather than being itself a moral calculator. They argue that, because people find incest and homosexuality disgusting, they are unlikely to engage in these behaviors anyway, and thus have little to lose by supporting norms against those behaviors. Moreover, because people perceive individuals engaging in disgusting behavior as being of low social value, they would condemn these behaviors to coordinate the collective exploitation of these vulnerable individuals (sect. R3.3).

We are not sure that these accounts escape the problems of other disgust-based theories. First, many behaviors are disgusting without being immoral, such as picking your nose and eating it in private (Kayyal, Pochedly, McCarthy, & Russell, 2015; Piazza et al., 2018; Royzman, Leeman, & Baron, 2009). If people have much to gain and little to lose by supporting norms against disgusting behavior, why does finding a behavior disgusting sometimes, but not systematically, lead to moral condemnation? Tybur & Lieberman's theory of incest and homosexuality, it seems to us, should specify why condemnation fixates on those disgusting behaviors and not others. Second, it predicts that disgust should robustly predict condemnations of incest and homosexuality. Yet correlations between disgust-sensitivity and condemnation of incest (Royzman et al., 2015), homosexuality (Schein, Ritter, & Gray, 2016), and other disgusting behaviors (Gray & Schein, 2016; Royzman et al., 2009; Schein et al., 2016) don't seem robust to controls nor to reflect a specific effect of disgust on moral condemnation (Landy & Piazza, 2017; Piazza et al., 2018).

R2.3.2. Guilt and shame are involved in, but not specific to puritanical morality

Zhu & Liu rightly point out that our theory was incomplete at the level of moral emotions. To fill this gap, they propose that guilt and shame characteristically underlie puritanical morality, in line with evidence that people feel guilt and shame after gambling, binge eating, masturbating, and failing to exercise. We fully agree that violations of puritanical norms can trigger guilt and shame. These emotions, however, are not *specific* to puritanical morality.

People feel guilt not only after gluttony, but also after lying, stealing, betraying, or cheating their partner – in line with remarks by Zhu & Liu. This is because guilt functions to compensate partners after violating contracts to restore reciprocal cooperation and your moral reputation (Fitouchi et al., *in press*). Because this adaptive challenge prevails across domains of reciprocal interaction, guilt is triggered by moral violations across domains.

R3. Puritanism and self-interest: Cooperation or strategic moralizing?

Many commentators view our theory as “idealistic” (Szocik) and “overly credulous” (Tybur & Lieberman). By overemphasizing cooperation, they argue, our account overlooks patriarchal coercion, cynical power struggles, and cruel punishments at the center of puritanism (Barenthin; DeScioli & Kurzban; Tybur & Lieberman; Szocik; Moon). In this section, we clarify the interplay between the puritanical moral contract and patriarchal coercion (sect. R3.1), moralistic punishment (sect. R3.2), and the strategic promotion of self-serving norms (sect. R3.3).

R3.1. Puritanical morality is not (only) patriarchal coercion

Several commentators argue that puritanical norms do not serve cooperation but are crafted by men to exclude women from public life and control their bodies and sexuality (Szocik; Barenthin). Across cultures, they note, people restrict sexuality more tightly for females than for males (Szocik; Royzman & Borislow; see also Barenthin; Weinstein & Baldwin). If puritanism serves to promote mutually beneficial cooperation, then “why hasn’t male sexual behavior been equally regulated throughout history?” (Szocik, para. 6).

Royzman & Borislow amplify the objection. Not only does our theory fail to explain this double standard, it even seems to predict the opposite. Compared to females, males have stronger sex drives and are more prone to antisocial behaviors when seeking sexual gratification (Buss, 2021). Thus, if puritanical restrictions aim to prevent antisocial effects of sexual self-control failures, they should regulate sex more strongly for males than for females, not the other way around! Let us answer these compelling arguments in several steps.

First, we fully agree that patriarchal coercion and objectification of women explain many restrictions on female sexuality. Our target article never denied this. We explicitly wrote that men’s interest to police women’s sexuality “surely underlies many sexual restrictions ... and is consistent with the frequent double standard favoring men in the moralization of sexuality” (target article, sect. 2.2). We also highlighted the “often-patriarchal nature of adultery proscriptions, which often sanction female’s infidelity more strongly than male adultery..., treating women as the property of their husband, father, or brothers” (target article, sect. 1.2, endnote 1).

However, our explanatory target – puritanical morality – differs from these patriarchal restrictions (sect. R1.3). Puritanical values refer to a specific type of sexual morality that prescribes sexual abstinence as just one facet of asceticism in general for both men and women. Condemnations of lustful marital sex and masturbation targeted both husbands and wives, boys and girls (Dabhoiwala, 2012; Seidman, 1990), and the historical Puritans moralized immodest clothing for both men and women (Bremer, 2009, p. 51). To convince you that puritanical asceticism is not inherently male-biased, take concrete data (Fig. R4). McIntosh

(2002) examined the proceedings of public courts from 255 villages and small towns in England (1370–1600), which sought to control quarrels, sexual misdeeds, and unruly alehouses. She notes that “the terms used to characterize all these offenses suggest that they violated...both self-control over one’s own behavior and the discipline that should be exercised by people in authority over their charges” (McIntosh, 2002, p. 68). Yet she finds that:

the majority of the lesser English courts that reported sexual problems presented both men and women at the same time...It seems, therefore, that local jurors were not concerned principally with female sexuality but rather were attempting to regulate disorderly sexual behavior wherever it occurred, among both men and women. (McIntosh, 2002, pp. 73–74; Fig. R4)

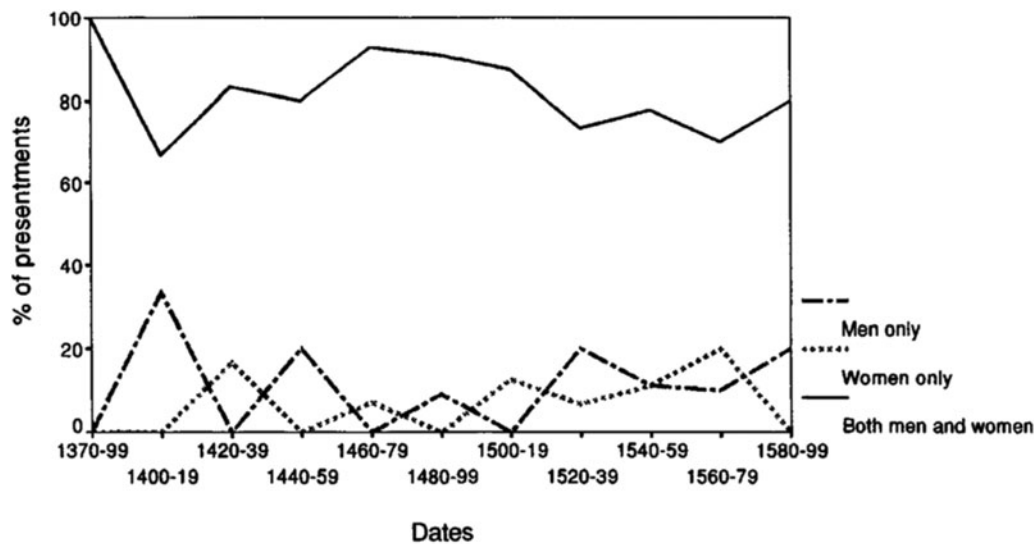
We agree, however, that puritanism is male-biased in other contexts. Szocik and Barenthin rightly note that female clothing, for example, is more tightly controlled across cultures. We agree that this double standard is largely rooted in patriarchal oppression. But note that – contrary to Royzman & Borislow’s suggestion – the moral disciplining model also explains this double standard. As Royzman & Borislow rightly note, men have stronger sex drives (Baumeister, Catanese, & Vohs, 2001), are more aroused by visual sexual stimuli (Hamann, Herman, Nolan, & Wallen, 2004), are more likely to engage in antisocial behaviors when desiring sex (Buss, 2021), and are perceived less able to control their sexual urges (Moon, Wongsomboon, & Sevi, 2021). Thus, exposing males to female sexual cues creates greater risk of antisocial behaviors than exposing females to male sexual cues. The moral disciplining model thus expects that, to prevent harmful self-control failures, modesty norms should regulate female clothing more strongly. Consistent with this view, the more people believe that men can’t control their sex drives, the more they moralize immodesty in women (Moon et al., 2021; target article, sect. 4.4).

R3.2. Moralistic punishment is for cooperation

DeScioli & Kurzban argue that, by overemphasizing cooperation, our account fails to explain another dark side of puritanism – harsh, moralistic punishments. We would explain why puritans prefer self-controlled people in partner choice, but not why they pay the costs of punishing impulsive people. Indeed, punishment is costly. As DeScioli & Kurzban note, this cost can be recouped if punishment disciplines partners for cooperation. But they ask, “why not simply look for a better partner instead of risking retaliation to try to teach a glutton self-control?”

Well, sometimes people can’t switch partners (Thomson et al., 2018). So they try to discipline the partners they have, provided they have enough bargaining power (Barclay & Raihani, 2016; von Rueden, Gurven, & Guala, 2012). If drunk neighbors disturb public peace and cause disorder and violence, what is less costly? Moving your whole family to another village – where there may be drug addicts anyway – or referring to the local court, where you’ll find other people keen to discipline drunkards? (see McIntosh, 2002).

The local court is key here. Contradicting DeScioli & Kurzban’s claim that “punishment is not for cooperation,” many communities organize to control free-riding for mutual benefit by sharing the costs of punishment (Hechter, 1988; Ostrom, 1990). They appoint accountable monitors to police for the common good and reward them with reciprocal payments



Graph 3.5 Courts that reported sexual misconduct: percentage of presentments by gender

Figure R4. Percentage of presentments for sexual misconducts to lesser English courts by gender, 1370–1599. Data and figure from McIntosh (2002).

or reputational benefits (Ostrom, 1990). Or they use coordinated sanctions to dilute the cost of punishment over multiple individuals who have a common interest in curbing free-riding (Boehm, 2012; Molleman, Kölle, Starmer, & Gächter, 2019). When people perceive undisciplined behaviors as threats to the public good (sect. R2.2.1), it's no mystery why they organize to punish them, just as they do for other forms of cheating (Ostrom, 1990).

The harshness of punishment is not specific to puritanism either. When people perceive others as untrustworthy, they see harsh, public punishments as necessary to sufficiently deter free-riding (Nettle & Saxe, 2021). They publicly execute traitors and cut off the hand of thieves. Consistent with this view, Fog cites evidence that tight social control is preferred in environments with high need for collective action. When people fear sex-addicts, gluttons, and drunkards in the same way, they want to control them as tightly as they do for any kind of cheaters. The fact that we, living in high trust and secure societies, see harsh punishments as pointless cruelties doesn't mean that people didn't see them as necessary evils to ensure the public good in less peaceful contexts. Reciprocal cooperation does not refer to anything that vaguely seems nice, but to interaction structures where people can achieve mutual benefit *if everyone refrains from exploiting others' cooperation*. The corollary is that exploiting others' cooperation makes you deserve punishment (André et al., 2022; Fitouchi & Singh, 2023).

R3.3. Strategic moralizing only exists because moral judgments encode mutual benefit in the first place

According to several commentators, puritanical morality arises from mechanisms evolved, not for mutually beneficial cooperation, but to promote moral norms advantageous to oneself (DeScioli & Kurzban; Moon; Tybur & Lieberman). Moon argues that people who are weaker and more vulnerable to disorder likely benefit from moralizing undisciplined behavior at the expense of people interested in freedom and creativity. DeScioli & Kurzban argue that puritanical morality stabilizes when

factions that benefit from these norms gain control over the rules (see also DeScioli, 2023). Tybur & Lieberman argue that puritanical moralizations serve to coordinate the exploitation of people of low social value (see sect. R2.3.1).

We agree that people often use moral judgments manipulatively (target article, sect. 2.2). However, we respectfully disagree with DeScioli & Kurzban, as well as Tybur & Lieberman, that moral condemnation can be reduced to a self-interested or coordination function. Moon notes that it's precisely because moral principles are widely about cooperation for mutual benefit that appealing to them is efficient to convince others to behave in accordance with your self-interest. In line with his remarks, we think the very reason why strategic moralizing is adaptive is that moral concepts encode information about mutual benefit in the first place. Let us explain.

Take Tybur & Lieberman's argument that moral condemnation serves to coordinate coalitional attack. To coordinate the exploitation of a low-value onanist, why would it be more efficient to say "Masturbation is wrong!" than to say "Let's all attack at 1 p.m. to get lots of benefits!?" The reason why moral language is useful is that it helps you justify your attack *in terms of the public good*, by presenting the exploited individual as a cheater who deserves punishment rather than an innocent victim of your selfish motives (Singh, 2021). Similarly, if you want to get others to drink less because you're vulnerable to disorder (Moon) or persuade them to adopt puritanical laws because they benefit you (DeScioli & Kurzban), why should it be more efficient to say "Drinking violates a moral duty!" than to say "Stop drinking! It's not in my interest!?" Again, the reason is that saying "Drinking is wrong" means "We would all be better off if everyone stopped drinking, not just me!" Strategic moralizing amounts to use moral arguments deceptively to convince others that a given behavior is mutually beneficial when in fact the behavior only benefits the condemner. But for people to benefit from using these manipulative signals, moral arguments must activate the notion of mutual benefit in the minds of receivers. Otherwise, recipients would have no interest in listening to such

arguments, and strategic moralizing would be of no use in the first place (see Dawkins & Krebs, 1978).

R4. Puritanism and folk-theories of self-control

R4.1. Puritanism depends on beliefs, not that asceticism signals self-control, but that it improves self-control

Contrary to what some commentators seem to have understood, we did not argue that bodily pleasures are moralized because they are perceived to *signal* low self-control and low cooperativeness (Curry & Sznycer; Giner-Sorolla & Myers; Graham et al.; Tierney, Cyrus-Lai, & Uhlmann [Tierney et al.]; Tybur & Lieberman). Without ascribing us this idea, other commentators themselves propose that purity violations are moralized because they signal impatience (Ellingsen & Mohlin), uncooperativeness (Murray et al.), or a low propensity to respect cultural norms in general (Giner-Sorolla & Myers).

These ideas based on signaling, however, are not sufficient to explain why people morally condemn bodily pleasures. Inferring that gluttons are impulsive or untrustworthy explain decisions to avoid gluttons on the cooperation market, but not the time and energy spent trying to reduce their gluttony – in line with remarks by Tybur & Lieberman and DeScioli & Kurzban. People in puritanical cultures don't just avoid impulsive individuals, they try to *stop them* from indulging by enacting legal prohibitions (Martin, 2009), reporting sinners to courts (McIntosh, 2002), and promoting techniques of self-discipline to help them curb carnal impulses (Bremer, 2009). This is why we argued that puritanical morality arises from beliefs, not that restraint *signals* self-control, but that restraint *strengthens* or *protects* self-control. This relates to Buchtel's fascinating data on the centrality of a "cultivated" character in Chinese lay concepts of morality. The core of puritanical morality is that people have a moral duty to cultivate character traits that will help them behave more cooperatively in the future (Fitouchi, André, & Baumard, 2022a; Fitouchi, André, Baumard, & Nettle, 2022b).

That said, we very much like Celniker, Ditto, Piff, & Shariff's (Celniker et al.) compelling insight that people may impose puritanical norms, not only to improve others' self-control, but also to test others' ability to control themselves in order to choose more disciplined partners. Consistent with this subtle idea, Kurdoglu compellingly argues that Turkish men use chastity norms in a fine-grained manner to extract information about potential partners' trustworthiness.

R4.2. When people don't believe that indulgence erodes self-control, they simply don't condemn it

Many commentators review evidence that indulgence in bodily pleasures, such as eating, drinking, and feasting is often used for social bonding, suggesting that people often perceive indulgence as facilitating cooperation rather than impeding it (Becker & Bernecker; Fu & Viera; Giner-Sorolla & Myers; Syme). Some of them see this as challenging our account, by showing that many societies have more positive views of bodily pleasure than our theory would assume (Fu & Viera; Giner-Sorolla & Myers, see also Becker & Bernecker).

We do not claim, however, that folk-theories that bodily pleasures threaten cooperation are universal. We argue that *when* people hold these folk-theories, they *morally condemn* bodily pleasures. This does not entail that humans everywhere should

exhibit these folk-theories. Many people don't morally condemn bodily pleasures, and our theory *predicts* that they should not hold these folk-theories. In fact, this prediction is supported by evidence cited by Fu & Viera, as well as Giner-Sorolla & Myers, that people reject puritanical norms when they perceive indulgence to promote cooperation, and prefer abstinence only when "self-control failure [is] seen as more harmful than innocuous" (Giner-Sorolla & Myers).

Starmans offers a fascinating way to test this relationship between folk-theories and moral judgments. She notes that beliefs that self-control can be trained or eroded likely emerge late in childhood. Thus, children should morally judge bodily pleasures differently than adults, to the extent that have different folk-theories of self-control (see also Starmans & Bloom, 2016). We fully agree and would very much like to see this prediction tested. We also agree with Syme that even puritanical adults likely hold more subtle folk-theories than those reviewed in the target article. In particular, allowing some periods of unrestrained indulgence – as opposed to continuous abstinence – may appear useful to better channel impulses in the rest of social life (Syme).

R4.3. On the (in)accuracy of puritanical folk-theories

R4.3.1. Why discipline others when disciplining is ineffective?

Our theory is agnostic about whether puritanical norms are objectively effective in improving self-control and cooperation (target article, sect. 3.4). But if puritanical norms don't work, Blancke asks, why would people try to discipline others in the first place?

Blancke builds on evidence that condemning a behavior makes you appear less likely to engage in it. Thus, condemning indulgence may signal that you are highly self-controlled, helping you attract cooperation partners. For the signal to be credible, however, receivers must ensure you're not a hypocrite. Paying the cost of disciplining others, Blancke argues, allows people to demonstrate their commitment to puritanical values and thus to credibly signal their self-control. This compelling idea nicely complements Celniker et al.'s argument that promoting puritanical norms provides benefits in partner choice (sect. R3.1).

Blancke's mechanism, however, seem to also require that people believe puritanical norms to improve self-control in the first place. To reap reputational benefits from moral condemnation, you must condemn behaviors that others see as deserving to be condemned. Unjustly condemning others for behaviors that nobody perceives as threatening the common good seems unlikely to bring reputational benefits – although this should be empirically tested. For people to gain reputational benefits from condemning bodily pleasure, then, others must at least find it plausible that unrestrained indulgence threatens the public good in some way.

R4.3.2. Societal implications of folk-psychological beliefs

Several commentators insist on the negative social consequences of puritanical folk-theories. We agree with Olivola that the intrinsic valorization of effort can lead to ineffective altruism, as well as with Becker & Bernecker that debunking puritanical theories of self-control – when they are false – allows preventing societies from missing the benefits of harmless pleasures. We particularly share Celniker et al.'s concern that limiting moralistic responses to welfare policies and liberal values requires understanding why moralizations of effort are so intuitive. Beyond self-control, understanding people's folk-theories of social phenomena often

seems crucial for policy design in many domains (Johnson & Nettle, 2020; Nettle & Saxe, 2021; Piff et al., 2020).

R5. Puritanism and moral variation

R5.1. Did puritanism fall in economically developed societies?

Many commentators note that puritanical morality is highly variable (Barenthin; Bonnefon; DeScioli & Kurzban; Fog; Syme). In our target article, we reviewed evidence that puritanical values have declined in western, educated, industrialized, rich, and democratic (WEIRD) societies (sect. 5). Several commentators question this idea (Olivola; Tierney et al.). In wonderful experiments, Tierney et al. found that, compared with Indian participants, Western participants judged needless work as less indicative of a good moral character in their deliberative responses, but not in their implicit responses. Tierney et al. thus suggest that WEIRD people are still puritanical on an intuitive level.

We are not sure that this is the case. Perceptions of a “good moral character,” measured by Tierney et al., seem to reflect perceptions of trustworthiness (Goodwin, 2015). Tierney et al.’s result, in other words, reflect that WEIRD participants still implicitly perceive people who engage in needless work as more trustworthy than people who don’t. This doesn’t really show, however, that WEIRD participants think that people have a *moral duty* to work even when they don’t need to – which would reflect a properly puritanical judgment. One possible explanation of the discrepancy between Tierney et al.’s implicit and deliberative measures, in fact, might be that WEIRD participants might have inhibited their intuitive distrust of the idle target in their deliberative response, precisely because they don’t have the moral intuition that the idle target has done anything wrong. These are of course only speculations, which we hope make some sense to Tierney et al.

This relates to Olivola’s suggestion that WEIRD societies would not have abandoned puritanism because they still value effort and find effortful actions more meaningful than easy ones (see Bloom, 2021; Inzlicht, Shenhav, & Olivola, 2018). Again, “valuing” effort in such general terms is not the same as viewing lack of effort as morally wrong behavior. Admiring mountaineers and feeling that effort adds meaning to one’s life is not the same as thinking that people have a *moral duty* to regularly engage in effortful activities – and that they ought to be punished if they don’t.

R5.2. Moral variation from universal computations

Echoing several commentaries (DiMaggio et al.; Fog; Kurdoglu), a key point of our theory is that the cultural variation of morality is not contradictory with moral judgments being produced by a universal computational procedure.

Moral variation, we argue, does not result from unconstrained coordination on arbitrary norms (DeScioli & Kurzban), nor from a plurality of moral cognitive systems (Graham et al.), but from flexible computations of mutual benefit based on variable inputs (sect. R2). Unlike Veit & Browning, we don’t think that puritanical morality in itself has “deep evolutionary roots” because it would have facilitated the self-control required by cooperative foraging in human evolution. The reason is that, as Syme rightly highlights, puritanical norms seem weak, if not absent, in many small-scale societies (target article, sect. 6.2). We did not argue that puritanical morality emerges from “innate intuitions about self-control” (DeScioli & Kurzban), which would be “adaptive”

in themselves (Becker & Bernecker). Rather, puritanical moral judgments arise from the regular computations of moral cognition – “What would be mutually beneficial if everyone did it?” (André et al., 2022; Levine et al., 2020) – when the latter are placed in environmental conditions where the costs of ascetic restraint appear worth the mutual benefits of social peace (sect. R2.2.1).

In line with this notion of cost–benefit balance (sect. R2), Bonnefon makes the provocative suggestion that, with the advent of autonomous cars and other machines, a new form of puritanical morality may replace the costly and effortful puritanism that WEIRD societies have abandoned. In order to prevent self-control failures, people will calculate that they are morally obliged, not to engage in laborious trainings of self-control, but simply to cede their agency to self-controlled machines (e.g., autonomous cars) that are never impulsive nor drunk – a much less costly way of protecting the public good from human impulses. This conjecture beautifully illustrates the point that morality is not a set of rigid rules, but the product of open-ended computations.

Financial support. This research was funded by Agence Nationale pour la Recherche (ANR-17-EURE-0017, ANR-10-IDEX-0001-02).

Competing interest. None.

Note

1. Other commentators argue that contraception (DeScioli & Kurzban) and homosexuality (DeScioli & Kurzban; Giner-Sorolla & Myers; Tybur & Lieberman) are moralized for reasons unrelated to self-control. We are agnostic on this question. But these behaviors are less remote from self-control than they seem. Although homosexuality was largely tolerated until late antiquity, its moralization rose as part of “a broad trend toward asceticism in the Hellenistic and late Roman empires,” which was “hostile to all forms of sexual pleasure, including homosexuality” (Greenberg & Bystry, 1982, pp. 517–520; see also Gaca, 2003). Like masturbation and unrestrained sex within marriage, homosexuality and contraception involve having sex only for *pleasure* (because they cannot lead to procreation) and thus to give free rein to hedonic consumption – the greatest fear of puritanical moralizers (Seidman, 1990).

References

- André, J.-B., Fitouchi, L., Debove, S., & Baumard, N. (2022). *An evolutionary contractualist theory of morality*. PsyArXiv. <https://doi.org/10.31234/osf.io/2hxgu>
- Barclay, P., & Raihani, N. (2016). Partner choice versus punishment in human Prisoner’s Dilemmas. *Evolution and Human Behavior*, 37(4), 263–271. <https://doi.org/10.1016/j.evolhumbehav.2015.12.004>
- Baumard, N., André, J.-B., & Sperber, D. (2013). A mutualistic approach to morality: The evolution of fairness by partner choice. *Behavioral and Brain Sciences*, 36(1), 59–78. <https://doi.org/10.1017/S0140525X11002202>
- Baumeister, R. F., Catanese, K. R., & Vohs, K. D. (2001). Is there a gender difference in strength of sex drive? Theoretical views, conceptual distinctions, and a review of relevant evidence. *Personality and Social Psychology Review*, 5(3), 242–273. https://doi.org/10.1207/S15327957PSPR0503_5
- Bloom, P. (2021). *The sweet spot: The pleasures of suffering and the search for meaning*. HarperCollins.
- Boehm, C. (2012). *Moral origins: The evolution of virtue, altruism, and shame*. Basic Books.
- Boyer, P., Firat, R., & van Leeuwen, F. (2015). Safety, threat, and stress in intergroup relations: A coalitional index model. *Perspectives on Psychological Science*, 10(4), 434–450.
- Bremer, F. J. (2009). *Puritanism: A very short introduction*. Oxford University Press.
- Buss, D. M. (2021). *Bad men: The hidden roots of sexual deception, harassment and assault*. Hachette UK.
- Clifford, S., Iyengar, V., Cabeza, R., & Sinnott-Armstrong, W. (2015). Moral foundations vignettes: A standardized stimulus database of scenarios based on moral foundations theory. *Behavior Research Methods*, 47(4), 1178–1198. <https://doi.org/10.3758/s13428-014-0551-2>
- Crone, D. (2022). *Conceptual issues with the moral foundation of purity: The case of religion*. PsyArXiv. <https://doi.org/10.31234/osf.io/3e8bv>

- Curry, O. S. (2016). Morality as cooperation: A problem-centred approach. In T. K. Shackelford & R. D. Hansen (Eds.), *The evolution of morality* (pp. 27–51). Springer. https://doi.org/10.1007/978-3-319-19671-8_2
- Dabhoiwala, F. (2012). *The origins of sex: A history of the first sexual revolution*. Oxford University Press.
- Dawkins, R., & Krebs, J. R. (1978). Animal signals: Information or manipulation? In J. R. Krebs & N. B. Davies (Eds.), *Behavioural ecology: An evolutionary approach* (pp. 282–309). Blackwell Scientific.
- DeScioli, P. (2023). On the origin of laws by natural selection. *Evolution and Human Behavior*, S1090513823000041. <https://doi.org/10.1016/j.evolhumbehav.2023.01.004>
- DeScioli, P., & Kurzban, R. (2009). Mysteries of morality. *Cognition*, 112(2), 281–299. <https://doi.org/10.1016/j.cognition.2009.05.008>
- Dungan, J., Chakroff, A., & Young, L. (2017). The relevance of moral norms in distinct relational contexts: Purity versus harm norms regulate self-directed actions. *PLoS ONE*, 12, e0173405. <https://doi.org/10.1371/journal.pone.0173405>
- Fitouchi, L., André, J.-B., & Baumard, N. (2022a). From supernatural punishment to big gods to puritanical religions: Clarifying explanatory targets in the rise of moralizing religions. *Religion, Brain & Behavior*, 13(2), 195–199. <https://doi.org/10.1080/2153599X.2022.2065352>
- Fitouchi, L., André, J.-B., & Baumard, N. (in press). Are there really so many moral emotions? Carving morality at its functional joints. In L. Al-Shawaf & T. K. Shackelford (Eds.), *The Oxford handbook of evolution and the emotions*. Oxford University Press.
- Fitouchi, L., André, J.-B., Baumard, N., & Nettle, D. (2022b). *Harmless bodily pleasures are moralized because they are perceived as reducing self-control and cooperativeness*. PsyArXiv. <https://doi.org/10.31234/osf.io/fzv43>
- Fitouchi, L., & Singh, M. (2023). Punitive justice serves to restore reciprocal cooperation in three small-scale societies. *Evolution and Human Behavior*. <https://doi.org/10.1016/j.evolhumbehav.2023.03.001>
- Gaca, K. L. (2003). *The making of fornication: Eros, ethics, and political reform in Greek philosophy and early Christianity*. University of California Press.
- Glowacki, L., & von Rueden, C. (2015). Leadership solves collective action problems in small-scale societies. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370, 1–13. <https://doi.org/10.1098/rstb.2015.0010>
- Goenka, S., & Thomas, M. (2022). When is sensory consumption immoral? *Journal of Personality and Social Psychology*. <https://doi.org/10.1037/pspp0000450>
- Goodwin, G. P. (2015). Moral character in person perception. *Current Directions in Psychological Science*, 24(1), 38–44. <https://doi.org/10.1177/0963721414550709>
- Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., & Ditto, P. H. (2013). Moral foundations theory: The pragmatic validity of moral pluralism. In P. Devine & A. Plant (Eds.), *Advances in experimental social psychology* (Vol. 47, pp. 55–130). Elsevier.
- Graham, J., Haidt, J., Motyl, M., Meindl, P., Iskiwicz, C., & Mooijman, M. (2018). Moral foundations theory. In K. Gray & J. Graham (Eds.), *Atlas of moral psychology* (pp. 211–223). Guilford.
- Gray, K., DiMaggio, N., Schein, C., & Kachanoff, F. (2022). The problem of purity in moral psychology. *Personality and Social Psychology Review*. <https://doi.org/10.1177/10888683221124741>
- Gray, K., & Keeney, J. E. (2015). Impure or just weird? Scenario sampling bias raises questions about the foundation of morality. *Social Psychological and Personality Science*, 6(8), 859–868. <https://doi.org/10.1177/1948550615592241>
- Gray, K., & Schein, C. (2016). No absolutism here: Harm predicts moral judgment 30x better than disgust – Commentary on Scott, Inbar, & Rozin (2016). *Perspectives on Psychological Science*, 11(3), 325–329. <https://doi.org/10.1177/1745691616635598>
- Greenberg, D. F., & Bystry, M. H. (1982). Christian intolerance of homosexuality. *American Journal of Sociology*, 88(3), 515–548.
- Hagen, E. H., & Garfield, Z. (2019). *Leadership and prestige, mothering, sexual selection, and encephalization: The computational services model* [preprint]. Open Science Framework. <https://doi.org/10.31219/osf.io/9bcdk>
- Haidt, J. (2012). *The righteous mind: Why good people are divided by politics and religion*. Knopf Doubleday.
- Hall, D. D. (2012). *A reforming people: Puritanism and the transformation of public life in New England*. UNC Press Books.
- Hamann, S., Herman, R. A., Nolan, C. L., & Wallen, K. (2004). Men and women differ in amygdala response to visual sexual stimuli. *Nature Neuroscience*, 7(4), 411–416. <https://doi.org/10.1038/nn1208>
- Hechter, M. (1988). *Principles of group solidarity* (1. paperback printing). University of California Press.
- Inzlicht, M., Shenav, A., & Olivola, C. (2018). The effort paradox: Effort is both costly and valued. *Trends in Cognitive Sciences*, 22, 337–349. <https://doi.org/10.1016/j.tics.2018.01.007>
- Johnson, E., & Nettle, D. (2020). Fairness, generosity and conditionality in the welfare system: The case of UK disability benefits. *Global Discourse*. <https://doi.org/10.1332/204378920X15989751152011>
- Kayyal, M. H., Pochedly, J., McCarthy, A., & Russell, J. A. (2015). On the limits of the relation of disgust to judgments of immorality. *Frontiers in Psychology*, 6, 1–9. <https://doi.org/10.3389/fpsyg.2015.00951>
- Kollareth, D., Brownell, H., Duran, J. I., & Russell, J. A. (2022). Is purity a distinct and homogeneous domain in moral psychology? *Journal of Experimental Psychology: General*, 152(1), 211–235.
- Landy, J., & Piazza, J. (2017). Reevaluating moral disgust: Sensitivity to many affective states predicts extremity in many evaluative judgments. *Social Psychological and Personality Science*, 10, 194855061773611. <https://doi.org/10.1177/1948550617736110>
- Le Goff, J. (1984). *Le refus du plaisir in L'amour et la sexualité: Vol. Amour et Sexualité en Occident* (pp. 52–59). Points Histoire.
- Levine, S., Kleiman-Weiner, M., Schulz, L., Tenenbaum, J., & Cushman, F. (2020). The logic of universalization guides moral judgment. *Proceedings of the National Academy of Sciences*, 117(42), 26158–26169. <https://doi.org/10.1073/pnas.2014505117>
- Martin, A. L. (2009). *Alcohol, violence, and disorder in traditional Europe*. Truman State University Press.
- McIntosh, M. K. (2002). *Controlling Misbehavior in England, 1370–1600*. Cambridge University Press.
- Merrill, L. T. (1945). The puritan policeman. *American Sociological Review*, 10(6), 766–776. <https://doi.org/10.2307/2085847>
- Molleman, L., Kölle, F., Starmer, C., & Gächter, S. (2019). People prefer coordinated punishment in cooperative interactions. *Nature Human Behaviour*, 3(11), 1145–1153. <https://doi.org/10.1038/s41562-019-0707-2>
- Mooijman, M., Meindl, P., Oyserman, D., Monterosso, J., Dehghani, M., Doris, J. M., & Graham, J. (2018). Resisting temptation for the good of the group: Binding moral values and the moralization of self-control. *Journal of Personality and Social Psychology*, 115(3), 585–599. <https://doi.org/10.1037/pspp0000149>
- Moon, J. W., Wongsomboon, V., & Sevi, B. (2021). Beliefs about men's sexual self-control predict attitudes toward women's immodest clothing and public breastfeeding [preprint]. PsyArXiv. <https://doi.org/10.31234/osf.io/67vh9>
- Nettle, D., & Saxe, R. (2021). “If men were angels, no government would be necessary”: The intuitive theory of social motivation and preference for authoritarian leaders. *Collabra: Psychology*, 7(1), 28105. <https://doi.org/10.1525/collabra.28105>
- Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action*. Cambridge University Press.
- Piazza, J., Landy, J. F., Chakroff, A., Young, L., & Wasserman, E. (2018). What disgust does and does not do for moral cognition. In N. Strohminger & V. Kumar (Eds.), *The moral psychology of disgust* (pp. 53–81). Rowman & Littlefield.
- Pietraszewski, D., Curry, O. S., Petersen, M. B., Cosmides, L., & Tooby, J. (2015). Constituents of political cognition: Race, party politics, and the alliance detection system. *Cognition*, 140, 24–39. <https://doi.org/10.1016/j.cognition.2015.03.007>
- Piff, P. K., Wiwad, D., Robinson, A. R., Aknin, L. B., Mercier, B., & Shariff, A. (2020). Shifting attributions for poverty motivates opposition to inequality and enhances egalitarianism. *Nature Human Behaviour*, 4(5), Article 5. <https://doi.org/10.1038/s41562-020-0835-8>
- Price, M. E., & Van Vugt, M. (2015). The service-for-prestige theory of leader-follower relations: A review of the evolutionary psychology and anthropology literatures. In S. M. Colarelli & R. D. Arvey (Eds.), *Biological foundations of organizational behavior* (pp. 397–477). University of Chicago Press.
- Royzman, E. B., & Borislow, S. H. (2022). The puzzle of wrongless harms: Some potential concerns for dyadic morality and related accounts. *Cognition*, 220, 104980. <https://doi.org/10.1016/j.cognition.2021.104980>
- Royzman, E. B., Kim, K., & Leeman, R. F. (2015). The curious tale of Julie and Mark: Unraveling the moral dumbfounding effect. *Judgment and Decision Making*, 10(4), 296.
- Royzman, E. B., Leeman, R. F., & Baron, J. (2009). Unsentimental ethics: Towards a content-specific account of the moral-conventional distinction. *Cognition*, 112(1), 159–174. <https://doi.org/10.1016/j.cognition.2009.04.004>
- Saxe, R. (2022). Perceiving and pursuing legitimate power. *Trends in Cognitive Sciences*, 26(12), 1062–1063. <https://doi.org/10.1016/j.tics.2022.08.008>
- Scanlon, T. (2000). *What we owe to each other*. Belknap Press.
- Schein, C., Ritter, R. S., & Gray, K. (2016). Harm mediates the disgust–immorality link. *Emotion*, 16(6), 862–876. <https://doi.org/10.1037/emo0000167>
- Seidman, S. (1990). The power of desire and the danger of pleasure: Victorian sexuality reconsidered. *Journal of Social History*, 24(1), 47–67. <https://doi.org/10.1353/jsh/24.1.47>
- Shweder, R. A. (2012). Relativism and universalism. In D. Fassin (Ed.), *A companion to moral anthropology* (pp. 85–102). Wiley.
- Singh, M. (2021). Magic, explanations, and evil: The origins and design of witches and sorcerers. *Current Anthropology*, 62(1), 2–29. <https://doi.org/10.1086/713111>
- Smith, J. M., & Parker, G. A. (1976). The logic of asymmetric contests. *Animal Behaviour*, 24(1), 159–175.
- Smith, K. M., & Kurzban, R. (2019). Morality is not always good. *Current Anthropology*, 60(1), 61–62.
- Starman, C., & Bloom, P. (2016). When the spirit is willing, but the flesh is weak: Developmental differences in judgments about inner moral conflict. *Psychological Science*, 27(11), 1498–1506. <https://doi.org/10.1177/0956797616665813>
- Suiming, P. (1998). The move toward spiritual asceticism in Chinese sexual culture. *Chinese Sociology & Anthropology*, 31(1), 14–24. <https://doi.org/10.2753/CSA0009-4625310114>

- Thomson, R., Yuki, M., Talhelm, T., Schug, J., Kito, M., Ayanian, A. H., ... Visserman, M. L. (2018). Relational mobility predicts social behaviors in 39 countries and is tied to historical farming and threat. *Proceedings of the National Academy of Sciences*, 115 (29), 7521–7526. <https://doi.org/10.1073/pnas.1713191115>
- Tomasello, M. (2020). The moral psychology of obligation. *Behavioral and Brain Sciences*, 43, e56. <https://doi.org/10.1017/S0140525X19001742>
- Tooby, J., Cosmides, L., & Price, M. (2006). Cognitive adaptations for N-person exchange: The evolutionary roots of organizational behavior. *Managerial and Decision Economics: MDE*, 27, 103–129. <https://doi.org/10.1002/mde.1287>
- Tyler, T. R. (2006). Psychological perspectives on legitimacy and legitimation. *Annual Review of Psychology*, 57(1), 375–400. <https://doi.org/10.1146/annurev.psych.57.102904.190038>
- Van Vugt, M., Jepson, S. F., Hart, C. M., & De Cremer, D. (2004). Autocratic leadership in social dilemmas: A threat to group stability. *Journal of Experimental Social Psychology*, 40(1), 1–13. [https://doi.org/10.1016/S0022-1031\(03\)00061-1](https://doi.org/10.1016/S0022-1031(03)00061-1)
- von Rueden, C. R., Gurven, M., & Guala, F. (2012). When the strong punish: Why net costs of punishment are often negligible. *Behavioral and Brain Sciences*, 35(1), 43.