Harmful transgressions qua moral transgressions: A deflationary view

Paulo Sousa[1*]

and

Jared Piazza[2]

(Manuscript accepted for publication, Journal *Thinking & Reasoning*)

[1]Institute of Cognition & Culture, Queen's University, Belfast, UK

[2]Department of Psychology, University of Pennsylvania, Philadelphia, USA

Correspondence should be addressed to Paulo Sousa, Institute of Cognition and Culture, Queen's University, Belfast, BT71NN, UK. Email: p.sousa@qub.ac.uk. The authors would like to thank Valerie Thompson, Katinka Quintelier and three anonymous reviewers for their thoughtful comments on the manuscript.

Abstract

One important issue in moral psychology concerns the proper characterization of the folk understanding of the relationship between harmful transgressions and moral transgressions. Psychologist Elliot Turiel and associates have claimed with a broad range of supporting evidence that harmful transgressions are understood as transgressions that are authority independent and general in scope, which, according to them, characterizes these transgressions as moral transgressions. Recently, many researchers questioned the position advocated by the Turiel tradition with some new evidence. We entered this debate proposing an original, deflationary view in which perceptions of basic-rights violation and injustice are fundamental for the folk understanding of harmful transgressions as moral transgressions in Turiel's sense. In this article, we elaborate and refine our deflationary view, while reviewing the debate, addressing various criticisms raised against our perspective, showing how our perspective explains the existent evidence, and suggesting new lines of inquiry.

Keywords: harm; moral transgression; punishment; moral psychology; social cognition

Harmful transgressions qua moral transgressions: A deflationary view

How do ordinary people understand the relationship between harmful transgressions and moral transgressions? Over the years, Turiel and associates (hereon the "Turiel tradition") have characterized moral transgressions (related to issues concerning harm, injustice *or* rights violations) in contraposition to conventional transgressions (related to issues concerning tradition and social coordination) by claiming that while the former are seen as authority independent and general in scope, the latter are seen as authority dependent and local in scope. Accordingly, they have claimed that harmful transgressions, as a subcategory of moral transgressions, are understood as authority independent and general in scope (see Nucci, 2001; Smetana, 1993; Tisak, 1995; Turiel, 1983, 2002).

There are two main debates regarding the Turiel tradition. First, researchers have argued that some conventional transgressions are seen as authority independent and general in scope and therefore that moral transgressions are not restricted to issues concerning harm, injustice *or* rights violations (see Haidt, Koller, & Dias, 1993; Haidt & Joseph, 2007; Nichols, 2002; Shweder, Mahapatra, & Miller, 1990). For example, Haidt et al. (1993) claimed that conventional transgressions such as a person privately masturbating with a dead chicken or cleaning a toilet with the national flag are moralized in Turiel's sense. Second, and more recently, researchers have argued that some harmful transgressions are not seen as authority independent and general in scope and therefore that Turiel's position on the folk understanding of harmful transgressions qua moral transgressions is unwarranted (see Kelly, Stich, Haley, Eng, & Fessler, 2007; Stich, Fessler, & Kelly, 2009; Quintelier, Fessler, & De Smet, 2012). For

example, Kelly et al. (2007) claimed that cases such as whipping as punishment and physical abuse as part of interrogation techniques are not moralized in Turiel's sense.

We entered the second debate raising both theoretical and methodological concerns: researchers, including Turiel and associates, have not been explicit enough about the hypothesis they are trying to advocate or question; the analysis of the related evidence has not been as nuanced as it needs to be. Moreover, based on these concerns, we have proposed an original, deflationary view on the issue, in which harmful transgressions are seen as authority independent and general in scope if the causation of harm is interpreted as involving basic-rights violation and injustice (Sousa, 2009a; Sousa, Holbrook, & Piazza, 2009; Piazza, Sousa, & Holbrook, 2013).

In this article, we elaborate and refine our deflationary view, while reviewing the second debate, addressing various criticisms raised against our perspective, showing how our perspective is supported by the existent evidence, and suggesting new lines of inquiry. Firstly, we characterize our deflationary hypothesis. Secondly, we discuss different ways of interpreting the hypothesis on the folk understanding of harmful transgressions qua moral transgressions and show that our deflationary hypothesis is the relevant one. Thirdly, we indicate how our hypothesis is supported by the broad evidence coming from the Turiel tradition. Then, we show in detail how we handle the more recent evidence, focusing on the case of punishment of a sailor who was drunk on duty by giving him five lashes with a whip, which has been the topic of much discussion. After that, we suggest new lines of inquire by discussing two further issues and envisaging possible amendments to our perspective (here, we also revisit the first debate described above). We conclude by summarizing and completing our discussion.

The Deflationary View

In order to characterize our deflationary perspective on the folk understanding of the relationship between harmful transgressions and moral transgressions we explicate our position concerning three aspects of this folk understanding.

Normativity and transgression. We assume that there are five basic deontic concepts, one of which being the concept of norm or normative (*cf.* Beller, 2008a, 2008b; Bender & Beller, 2003; Bucciarelli & Johnson-Laird, 2005). Not to put too fine a point on it, three of the basic deontic concepts are inter-defined as follows: (i) FORBIDDEN—what is not permitted to do; (ii) OBLIGATORY—what is permitted to do and forbidden not to do; (iii) PERMITTED—what is obligatory to do or what is neither obligatory nor forbidden to do. For the sake of illustration, the action of killing an innocent person is part of what is forbidden, the action of providing adequate care for one's children is part of what is obligatory, and the actions of providing adequate care for one's children (what is obligatory) and going to the cinema (what is neither obligatory nor prohibited) are part of what is permitted. The remaining deontic concepts relate to the previous ones as follows: (iv) NORMATIVE—what is obligatory to do or what is forbidden to do; (v) DISCRETIONARY—what is neither obligatory nor forbidden to do. Here, the actions of providing adequate care for one's children, which is obligatory, and of killing an innocent person, which is forbidden, illustrate the two types of norms that constitute what is normative, and that of going to the cinema illustrates what is discretionary.

What is discretionary entails a sense of what is neither right nor wrong; that is, doing or not doing what is discretionary is seen as a question of personal choice. What is normative entails a sense of what is right or a sense of what is wrong; that is, doing what is obligatory is seen as the right thing to do and doing what is forbidden is seen as the wrong thing to do. Furthermore,

because it is obligatory to omit what is forbidden (e.g., obligatory to not kill an innocent person) and is forbidden to omit what is obligatory (e.g., forbidden to not provide adequate care for one's children), both doing what is obligatory and omitting what is forbidden are seen as right, and both doing what is forbidden and omitting what is obligatory are seen as wrong. Thus, the sense of wrong comes from what transgresses the normative, that is, from actions and omissions that do not follow norms. Fig. 1 represents the concepts delineated so far.

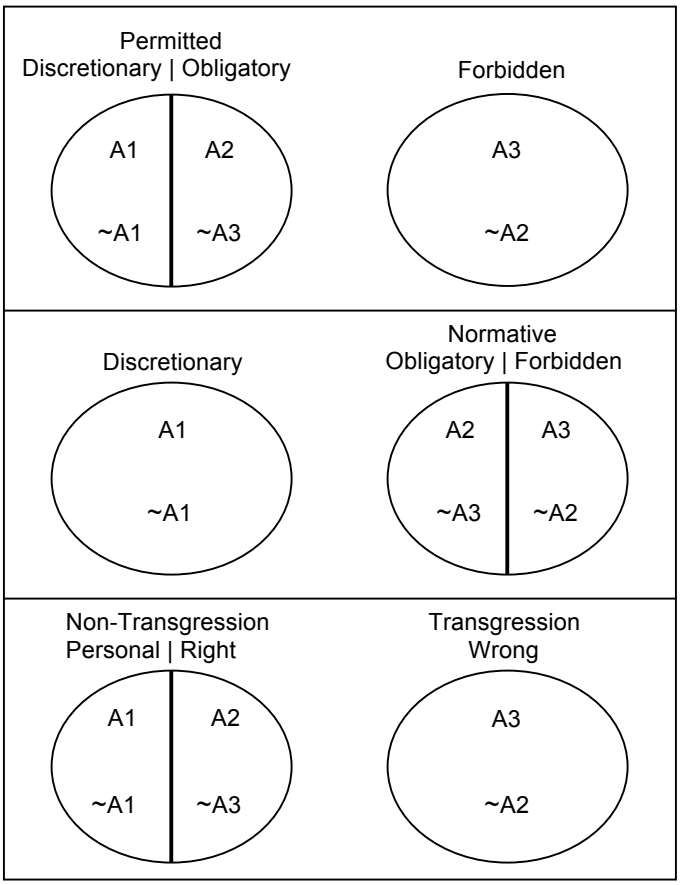

Figure 1. Deontic concepts (Adapted from Sousa 2009b).
A = Action; ~A = Omission of A; A1 = Discretionary actions;
A2 = Obligatory actions; A3 = Forbidden actions.

Moral norms and moral transgressions. The moral-psychology literature uses the word "moral" in two different senses (cf. Machery & Mallon, 2010). In one sense, the emphasis is on a specific type of normative content—e.g., moral norms forbid injustice; moral transgressions involve injustice. In another sense, the emphasis is on a specific type of normative conviction—moral norms involve a strong conviction that an action or omission is to be forbidden; moral transgressions are seen as unquestionably wrong (note that our focus is on what is forbidden/transgression/wrongness, not on what is obligatory/non-transgression/rightness—see Figure 1). In this article, we use "moral" only in the latter sense, and we characterize the normative conviction specifying moral norms/transgressions, such as the conviction that one should not kill an innocent person and that killing an innocent person is wrong, in contraposition to the normative conviction specifying conventional norms/transgressions, such as the conviction that one should eat a meal with utensils and that eating a meal with one's fingers is wrong. Following the Turiel tradition, we take this difference in convictions to involve the contrasting values of two variables:

Authority Contingency

Moral transgressions (and the norms forbidding them) are seen as independent of authority—their wrongness (and the normative force of the norms forbidding them) is assumed not to be cancelable by the decision of any authority.

Conventional transgressions (and the norms forbidding them) are seen to depend on authority—their wrongness (and the normative force of the norms forbidding them) is assumed to be cancelable by the decision of a legitimate authority.

Generalizability

Moral transgressions (and the norms forbidding them) are seen as general in scope—their wrongness (and the normative force of the norms forbidding them) is assumed to extend to different places and/or times.

Conventional transgressions (and the norms forbidding them) are not seen as general in scope—their wrongness (and the normative force of the norms forbidding them) is not assumed to extend to different places and/or times.

The notion of generality at stake here requires some clarification. First, the assumption that the wrongness of moral transgressions (and the normative force of the norms forbidding them) extends to different places and/or times is equivalent to the assumption that their wrongness (and normative force) is independent of any incongruous social consensus that may be deemed to exist in different places and/or times. For example, I would assume that keeping slaves is a transgression that is general in scope if and only if I were to judge that keeping slaves is still wrong in places and/or times where slavery is practiced and/or permitted. Second, the assumption that a transgression is general in scope *does not entail the assumption that people in different places or times share the same point of view*; the latter assumption corresponds to a different, logically independent notion of generality—namely, generality as an assumption that moral norms are universally shared. To illustrate, I could assume both that keeping slaves is morally forbidden and that people in a different time deemed slavery discretionary or obligatory. In this case, I would just think that these people had not yet acquired relevant normative knowledge, that is, that they did not know that keeping slaves is wrong.

Harmful transgressions and moral transgressions. We define harmful transgressions as those transgressions whose actions or omissions cause pain or suffering, whether this pain or

suffering is caused by physical or symbolic processes. We propose that harmful transgressions in this sense are categorized as transgressions whose wrongness is authority independent and general in scope if the causation of harm is interpreted as involving basic-rights violation/injustice—i.e., $\forall x ((Hx \& Ix) \rightarrow (MTx))$.[1] This is because we hypothesize that the folk concept of basic-rights violation/injustice implies transgression as well as authority independence and generality—i.e., $\forall x (Ix \rightarrow MTx)$.

There are three important points to clarify here. First, the expressions "basic-rights[2] violation" and "injustice", as we use them, comprise two different ways of referring to the same thing. Following some recent work on the evolution of cooperation and morality (see Baumard, Andre, & Sperber, 2013; Sperber & Baumard, 2012), we hypothesize that humans possess a specialized cognitive system naturally selected to deal with mutualistic interactions, a system that interprets these interactions as if they were part of a social contract entailing that individuals have some basic rights, such as the right to not be subjected to harm for selfish reasons, and have complementary obligations to not infringe the basic rights of others, such as the obligation to not harm other individuals for selfish reasons. Moreover, we hypothesize that, given this assumption of a social contract, any transgression of these complementary obligations is conceptualized as an instance of basic-rights violation *and* injustice—e.g., injuring another person for personal gain would be seen as an instance of basic-rights violation and injustice.

---

[1] "*H*", "*I*" and "*MT*" stand for the following properties, respectively: harmful, involving basic-rights violation/injustice, and moral transgression. "$\forall$" represents a universal quantifier, "$\rightarrow$" represents a conditional, and "x" is a variable ranging over actions and omissions. This formula should be read as: for any item x, if the item has property *H* and has property *I*, then it has property *MT*.

[2] The general concept of rights at stake here is the concept of "claim-rights" (for an analysis of this folk concept, see Jackendoff, 1999).

Before moving to the second point, it is important to note that we are not claiming that human beings are always seen as individuals having basic rights. Throughout history specific human beings have been denied the full status of person (i.e., an individual perceived as deserving moral standing—see Singer, 2011) and hence been denied basic rights. Also, when we talk about the assumption of a basic right to not be subjected to harm *for selfish reasons*, our claim is about how ordinary people would interpret the proximal reason for the causation of harm—whether this reason is deemed to take or not take into account the interest of the person being harmed. For example, when a doctor vaccinates a patient *in order to* cure the patient from a disease, people (including the patient) will interpret the causation of pain as in accordance with the interest of the patient (hence, not as motivated by a selfish reason). Whether people would interpret this proximal reason as an instrumental desire to satisfy another, more distal selfish desire of the doctor is irrelevant; more broadly, our claim is independent of whether human motivation is (or is perceived to be) *ultimately* guided by selfish motivations, such as to gain social prestige (see Stich, Doris, & Roedder, 2010). Now, if the doctor stabbed the patient with a sadistic motivation and without the intent to inject a vaccination, then people would attribute to her a selfish reason. Likewise, if the doctor injected the vaccination when there is available an equally viable, painless alternative, in order to receive some financial benefit for choosing the more painful procedure over the less painful one, people would attribute to her a selfish reason.

The second point concerns the basis for hypothesizing that the folk concept of basic-rights violation/injustice implies transgression as well as authority independence and generality. This folk concept implies transgression because, as we discussed above, it refers to actions and omissions that do not follow a complementary obligation, namely, the obligation not to infringe the basic rights of others. In other words, this implication is an immediate semantic inference

(see Figure 1). This folk concept implies wrongness that is authority independent and general in scope because it refers to actions or omissions that are incompatible with the logic of the social contract delineated by the specialized cognitive system naturally selected to deal with mutualistic interactions. In other words, this implication works like a *reductio ad absurdum* inference: it follows from the logic of the social contract that it is absurd (i.e., 'contradictory') for an authority or social consensus to cancel the wrongness of actions or omissions involving basic-rights violation/injustice; therefore, the wrongness of these actions or omissions is authority independent and general in scope. According to our perspective here, it is important to point out that, in terms of the distinction between explicit (i.e., stored) and tacit (i.e., not stored but inferable) assumptions (see, e.g., Lycan, 1986), ordinary people, including young children, need not assume explicitly that transgressions involving basic-rights violation/injustice are authority independent and general in scope; they may do so only if prompted to think about the matter, as in the studies we shall describe later on.

The third point concerns an issue raised by Stich et al. (2009, p. 96). Since we hypothesize that the folk concept of basic-rights violation/injustice implies transgression as well as authority independence and generality, all actions interpreted as involving basic-rights violation/injustice are to be understood as moral transgressions, independent of other properties, such as being harmful, that one may attribute to the action—i.e., $\forall x\ (Ix \rightarrow MTx)$. Thus, they say, the concept of harm does not play a central role in our perspective since this perspective tells "nothing distinctive about moral judgments that involve harmful acts".

Indeed, harm in itself does not play a fundamental conceptual role in our perspective, contrary to some recent interpretation of our work (see Waldmann, Nagel, & Wiegmann, 2012, p. 286, 287). In other words, our view on the folk understanding of the relation between harmful

transgressions and moral transgressions is deflationary concerning harm. However, we would not claim that there is nothing distinctive about harm in relation to moral transgressions. Without assuming too rosy a picture of human psychology, we accept that in many circumstances humans have aversive affective reactions to the perception of pain and suffering, and the performance of harmful acts (Cushman, Gray, Gaffey, & Mendes, 2012), and that these reactions may play a role in the cultural evolution and stabilization of norms by biasing normative systems to increasingly categorize harmful actions as involving basic-rights violation/injustice, and hence, as morally forbidden (cf. Nichols, 2004). Moreover, without assuming that "badness" implies wrongdoing, harmful transgressions often seem to be regarded as the most serious moral transgressions—i.e., there may be some relevant distinctiveness in how people perceive the badness of harmful transgressions involving basic-rights violation/injustice compared to those transgressions involving basic-rights violation/injustice without involving harm (e.g., ending the life of an innocent person painlessly via euthanizing drugs and without consent).

The Relevant Hypothesis

A major problem in the current debate on the folk understanding of the relationship between harm and morality is that researchers have not been explicit enough about the hypothesis they are trying to advocate or question. In this section, we briefly delineate the different hypotheses existent in the current debate. A hypothesis that is trivially false or tautological cannot be relevant in the context of the empirical sciences, and a hypothesis that does not portray a reasonably clear psychological model cannot be relevant in the context of the cognitive sciences. Accordingly, in this section we aim to show that, once the different hypotheses on the folk understanding of the relationship between harm and morality are explicated, it becomes apparent that while our deflationary hypothesis is relevant in these

respects, the other hypotheses are not.

Turiel and associates sometimes talk about harm as a moral domain separate from justice and rights, which implies that they are advocating that harmful actions themselves are seen as transgressions whose wrongness are general in scope and authority independent (see e.g., Turiel, 1983, pp. 39–40; Wainryb, 1991, pp. 842-843); sometimes they acknowledge that harmful actions themselves are not sufficient for moral wrongdoing because they are not sufficient for wrongdoing, but they do not delineate a hypothesis specifying what else is necessary beyond an analysis of the reason for the harmful action (see Turiel, 1983, p. 43). Thus, when the Turiel tradition claims that harm is seen as authority independent and general in scope, it is difficult to know what exactly is the hypothesis they are advocating.

Debating with the Turiel tradition, Quintelier et al. (2012, p. 193) claim to be trying to falsify the hypothesis, which they attribute to Turiel, that harmful actions themselves entail transgressions whose wrongness are general in scope and authority independent—i.e., $\forall x\,(Hx \rightarrow MTx)$. On the other hand, Stich et al. (2009, p. 94) claim that, given that Turiel and associates have not been clear, one should try to falsify the interpretation of Turiel's hypothesis that has been accepted by a number of influential psychologists and philosophers, which states that if an action is considered to be harmful, and if it is also considered to be a transgression even for reasons that have "little or nothing to do with the fact that" it is harmful, its wrongness is considered to be authority independent and general in scope—i.e., $\forall x\,((Hx \,\&\, Tx) \rightarrow (MTx))$.

It is important to emphasize that we use "harmful action" in the sense of *actions that cause pain or suffering*, which includes neither transgression nor rights violation/injustice. Both Turiel (1983) and Stich et al. (2009) use "harmful action" in this sense, but Quintelier et al. (2012) are not explicit about this. They may be using "harmful action" in the sense of *wrongful*

actions that cause pain or suffering, which includes transgression and is one of the ordinary meanings of "harm." If so, Quintelier et al. would be claiming that Turiel's hypothesis is tantamount to the second hypothesis characterized above—i.e., ∀x ((*H*x & *T*x) → (*MT*x)). It is also worth noticing that sometimes researchers even use "harm" in a third sense that includes both wrongdoing and rights violation/injustice: "Harm, broadly construed to include psychological harm, injustice, and violations of rights, may be important in the morality of all cultures" (Haidt et al., 1993, p. 613). Actually, some of the disagreement in the literature is due merely to the fact that authors sometimes oscillate between these different meanings of "harm."

Now, we would like to argue that, regardless of what Turiel and associates have had in mind, the hypothesis coming from our perspective should be the one at stake—i.e., ∀x ((*H*x & *I*x) → (*MT*x)). The first hypothesis above, ∀x (*H*x → *MT*x), is not worth testing because it is trivially false—there are too many cases of harmful actions that people consider to be permitted and even obligatory (e.g., causing harm as deserved punishment, medical treatment, self-defense, physical training, etc.), hence not as a case of transgression or moral transgression. On the other hand, the second hypothesis, ∀x ((*H*x & *T*x) → (*MT*x)), does not constitute a clear psychological model of the folk understanding involved—it does not specify what is supposed to establish that harmful actions are transgressions and are transgressions whose wrongness are authority independent and general in scope. Moreover, it is doubtful that this hypothesis corresponds to a widely accepted interpretation of Turiel's hypothesis. In support of this claim, Stich et al. (2009) refer only to the work of philosopher Shaun Nichols. Nevertheless, take these passages from Nichols:

(…) "Normative theory" is not intended in any inflated sense. Rather, even a motley set

> of rules prohibiting certain behaviors will count as a normative theory. (…) The normative theory of central interest to us, however, is *the normative theory prohibiting harmful actions*. The operative notion of harm also needs to be qualified. Unless otherwise noted, "harm" is restricted to psychological harms like pain and suffering.
>
> (…) Of course, this body of information about moral violations cannot be captured by a simple rule like "a behavior is wrong if it causes harm." At least among adults, behavior that is unintentionally harmful is often not regarded as transgressive. Sometimes a person can even intentionally cause suffering without incurring negative moral judgments. For example, applying an anti-infective to a child's scraped knee causes the child sharp pain, but we do not judge this to be morally wrong. *Among other things, the normative theory provides the basis for distinguishing wrongful harm from acceptable harm*. (Nichols, 2004, pp.16-17; the emphases are ours)

The first passage shows that, for Nichols, the wrongfulness of harmful actions is related to norms prohibiting harmful actions. Therefore, it is not that harmful actions could be considered an instance of wrongdoing even “for reasons that have little or nothing to do with the fact that they were harmful” (Stich et al. 2009, p. 94). It seems that even Nichols could not be plausibly interpreted as entertaining the hypothesis delineated by Stich et al. In other words, Stich et al. might be trying to falsify a hypothesis that no one advocates.

The second passage shows that, for Nichols, the normative theory includes some additional *conceptual* criterion specifying when a harmful action is an instance of wrongdoing, though he does not explicate what the content of this criterion could be. Our hypothesis is based

on the assumption that the folk concept of basic-rights violation/injustice implies transgression as well as authority independence and generality—i.e., ∀x ((*H*x & *I*x) → (*MT*x)) because ∀x (*I*x → *MT*x). It contains a sufficient conceptual criterion for distinguishing harmful actions that are seen as transgressions from those that are not. In addition, it contains a rationale for inferring that the wrongness of harmful transgressions is authority independent and general in scope. In sum, our hypothesis provides a reasonably clear psychological model.

Quintelier et al. (2012, pp. 193-4) argue on other grounds that our hypothesis could not be relevant. Because we assume that the folk concept of basic-rights violation/injustice implies transgression as well as authority independence and generality, our hypothesis, they say, becomes true by definition and hence unfalsifiable. It is indeed true that if one assumes that the concepts of basic-rights violation/injustice imply moral transgression logically one should assume that harmful actions involving basic-rights violation/injustice are moral transgressions. However, this does not make our perspective tautological in any relevant sense, since all our claims are hypotheses about folk understanding, and it is an empirical question whether humans in all cultures possess the concept of basic-rights violation/injustice we characterized, and whether this folk concept implies transgression as well as authority independence and generality. In other words, although "[∀x (*I*x → *MT*x)] → [∀x ((*H*x & *I*x) → (*MT*x))]" is logically true, our hypothesis is that "ordinary people assume ∀x ((*H*x & *I*x) → (*MT*x)) because they assume ∀x (*I*x → *MT*x)", which is an empirical hypothesis.

Evidence from the Turiel Tradition

Our perspective is supported by the broad evidence coming from the moral/conventional task, which is the methodology utilized to probe the difference between moral and conventional transgressions in the Turiel tradition (e.g., Helwig, Hildebrandt, & Turiel, 1990; Nucci, 2001;

Nucci & Turiel, 1978; Smetana, 1981; Smetana & Braeges, 1990; Turiel, 1983; Weston & Turiel, 1980; Yau & Smetana, 2003). In this task, participants are presented with scenarios of harmful actions motivated by selfish reasons without including additional reasons that may justify the causation of harm and, consequently, as actions to be interpreted as involving basic-rights violation/injustice (e.g., an innocent child is pushed off a swing by another child).[3] The harmful actions are described neither as transgressions nor as moral transgressions, but simply as something someone does. For each scenario wherein a harmful action A is described, a sequence of probes is posed:

(1) *Manipulation-check probe*: Is it OK for actor X to perform A? Yes No

(2) *Justification probe*: Please, explain your answer …

(3) *Authority-contingency probe*: Now, what if a legitimate authority says that it is OK to perform A. Would it be OK for actor X to perform A? Yes No

(4) *Generalizability probe*: In another place (and/or time), people think that it is OK to perform A. Is it OK for people in this other place (and/or time) to perform A? Yes No

The manipulation-check probe verifies whether participants indeed interpreted the harmful action as a transgression (whether they answered "No"—i.e., Not-OK), since the aim of the task is to probe whether harmful *transgressions* are seen as moral transgressions. The prediction related to the justification probe is that participants will explain the wrongness of the harmful action in terms of basic-rights violation or injustice. Finally, the prediction related to both the authority-contingency and generalizability probes is that participants will answer "No"

[3] In our discussion, we leave aside the conventional side of the moral/conventional task and we focus on harmful actions instead of the broader scope of possible immoral actions.

(i.e., Not-OK), indicating thereby that they judge the harmful transgression to be a moral transgression.

These predictions have been repeatedly borne out: from a young age and in different cultural contexts, the great majority of participants evince the predicted pattern of responses (see references above). These results can be explained by our perspective: participants' justifications (invoking basic-rights violation/injustice) indicate that participants consider the harmful actions to be transgressions because the causation of harm involves basic-rights violation/injustice, and, as we have argued, this also leads them to see the harmful transgressions as authority independent and general in scope.

Stich et al. (2009, p. 95-96) claim that, according to our perspective, the fact that people evince a No-No response pattern to instances of harmful actions in the moral/conventional task shows simply that their answers are rational in that they are applying the presumed concepts coherently—namely, $((H\mathrm{x}_1 \,\&\, I\mathrm{x}_1) \rightarrow (MT\mathrm{x}_1))$. As Stich et al. themselves emphasize, this demonstration of rationality in itself is not without interest, given the literature on reasoning and decision-making showing that on many tasks people do not conform to normative standards of rationality (for a review, see Samuels & Stich, 2004). However, we take this demonstration of rationality to have confirmatory significance too: participants' No-No response pattern (including the related justifications) not only indicates that participants are deploying their conceptual competence properly, it also provides good evidence that they possess the types of concepts we have outlined.

The Case of the Drunken Sailor

In the moral/conventional task as normally utilized by the Turiel tradition, participants are presented with harmful actions motivated by selfish reasons without including additional

reasons that may justify the causation of harm in order to isolate what is fundamental to harmful transgressions qua moral transgressions. These harmful actions, which we call cases of *simple harm*, should be broadly contrasted with harmful actions whose motivations are interpreted as including reasons that may justify the causation of harm, which we call cases of *complex harm* (see Sousa et al., 2009; Piazza et al., 2013; Turiel and associates have sometimes referred to these as "nonprototypical" cases; see Turiel, Hildebrandt, & Wainryb, 1991; Wainryb, 1991).

Cases of complex harm that are interpreted as motivated exclusively by reasons that justify the causation of harm are uniformly considered to be OK; some examples include causing harm out of self-defense (wherein the person is deemed to have the right to cause harm), as deserved punishment (wherein the person being harmed is deemed to have had her basic right suspended because of her wrongful actions), or as medical treatment (wherein the person being harmed is deemed to benefit from it). Cases of complex harm where the causation of harm is seen as totally justified are irrelevant as evidence for or against any perspective on the relationship between harmful transgressions and moral transgressions because they are not conceived to be transgressions at all. However, cases of complex harm often generate a fair amount of disagreement over whether the causation of harm is justified (e.g., see Piazza et al., 2013). This is because people may construe the motivating reasons differently, prioritize certain reasons over others, or have different criteria about what level or kind of harm is justifiable.

Using a methodology modeled after the moral/conventional task, a group of researchers has utilized scenarios involving cases of complex harm, such as whipping as punishment and physical abuse as part of military training or interrogation techniques, in order to provide evidence against hypotheses on the folk understanding of the relationship between harm and morality like the ones we discussed above in the section "The Relevant Hypothesis" (Kelly et al.,

2007; Stich et al., 2009; Quintelier et al., 2012). Here we review the case of punishing a sailor who was drunk on duty by giving him five lashes with a whip, showing how this case offers evidence in support of our hypothesis.[4]

Prima facie evidence. Kelly et al. (2007) and Sousa et al. (2009) presented participants with one of the following paired scenarios and corresponding questions (in Sousa et al., 2009, participants were also asked to justify their Yes or No answers):

Whipping-authority Pair

(1) Mr. Adams is an officer on a large modern American cargo ship in 2004. One night, while at sea, he finds a sailor drunk at a time when the sailor should have been monitoring the radar screen. After the sailor sobers up, Adams punishes the sailor by giving him 5 lashes with a whip.

Question: Is it OK for Mr. Adams to whip the sailor? Yes No

(2) Now suppose that the Captain of the modern cargo ship had told Mr. Adams that 'On this ship it is OK for officers to whip sailors'.

Question: Is it OK for Mr. Adams to whip the sailor? Yes No

Whipping-generality Pair

(1) Mr. Adams is an officer on a large modern American cargo ship in 2004. One night, while at sea, he finds a sailor drunk at a time when the sailor should have been monitoring the radar screen. After the sailor sobers up, Adams punishes the sailor by giving him 5 lashes with a whip.

Question: Is it OK for Mr. Adams to whip the sailor? Yes No

(2) Three hundred years ago, whipping was a common practice in most navies and on cargo ships. There were no laws against it, and almost everyone thought that whipping was an appropriate way to discipline sailors who disobeyed orders or were drunk on duty.

[4] We leave aside the military cases because they introduce other issues related to utilitarian harm that we discuss in detail elsewhere (see Piazza et al., 2013). It is worth pointing out that the above researchers utilized a slavery scenario as well, but this scenario in fact constitutes a case of simple harm, whose results, contrary to their claims, completely support positions like ours and Turiel's (see Sousa, 2009a; Sousa et al., 2009).

> Mr. Williams was an officer on a cargo ship 300 years ago. One night, while at sea, he found a sailor drunk at a time when the sailor should have been on watch. After the sailor sobered up, Williams punished the sailor by giving him 5 lashes with a whip.
>
> Question: Is it OK for Mr. Williams to whip the sailor? Yes No

The first scenario of each pair presented the action in a way that, supposedly, participants would deem it an instance of wrongdoing; hence, the question of the first scenario fulfilled the same role of the manipulation-check probe of the standard moral/conventional task. Each second scenario, together with its question, played a similar role to either the authority-contingency probe or the generalizability probe of the standard task.

For each pair, a participant could evince four response patterns: No-No, No-Yes, Yes-Yes, and Yes-No (the first Yes or No are related to the first scenario, that is, to the manipulation-check probe; the second Yes or No to the second scenario, that is, to either the authority-contingency probe or the generalizability probe). A No-No pattern is prima facie evidence confirming our perspective whereas a No-Yes pattern is prima facie evidence disconfirming it (or confirming the alternative hypothesis that harmful transgressions are not seen as moral transgressions). The Yes-Yes pattern is irrelevant to test our (or any) perspective on how the folk conceptualize harmful transgressions qua moral transgressions, because it indicates that the participant did not judge the harmful action to be a transgression in the first place. In these contexts, the Yes-No pattern not only is irrelevant but also indicates participants' error or misinterpretation of the scenario. Thus, we exclude it from our following discussion. The percentages of each response pattern in Kelly et al. and Sousa et al. studies are represented in Table 1, with the last two columns representing the revised percentages when irrelevant participants (Yes-Yes) are removed from the analysis.

Table 1

Percentage of participants with each response pattern with regard to the Whipping scenario (' = percentage when participants with the Yes-Yes pattern are eliminated)

Study	Paired Scenarios	*N*	No-No	No-Yes	Yes-Yes	No-No'	No-Yes'
Kelly et al. (2007)	Authority	195	76.5%	17.5%	6.0%	81.5%	18.5%
	Generality	198	49.0%	41.0%	10.0%	54.5%	45.5%
Sousa et al. (2009)	Authority	33	94.0%	3.0%	3.0%	97.0%	3.0%
	Generality	30	77.0%	16.0%	7.0%	82.0%	18.0%
Quintelier et al. (2012)	Generality	416	51.0%	16.0%	33.0%	76.0%	24.0%

The great majority of participants evinced the No-No response pattern in both studies, except for the whipping-generality pair in Kelly et al.'s study. However, as Frazer (2012) pointed out in discussing these results, there is an asymmetry between the two scenarios of the whipping-generality pair that can explain its high percentage in No-Yes answers. Many participants may have interpreted the situation three hundred years ago as much more dangerous than nowadays in that they may have envisioned that at that time there was a threat of piracy. If so, being drunk on duty three hundred years ago indicates greater recklessness and culpability than nowadays, which leads one to see an extreme form of punishment as justified and therefore as OK in the second scenario. This interpretation is corroborated by the results of a more recent whipping-generality study devised to eliminate the piracy-confound problem (see Quintelier et al., 2012).[5]

[5] Quintelier et al.'s study had a second aim as well—to demonstrate a methodological problem with current phrasings of the generalizability probe in the moral/conventional task in the Turiel tradition. We disagree with their representation of the Turiel Tradition with regards to this issue, but to elaborate on this would take us outside the

Participants were presented with the following paired scenarios and corresponding questions (participants were also asked to justify their Yes or No answers):

Whipping-generality Pair

(1) Mr. Johnson is an officer on a cargo ship in 2010, carrying goods along the Atlantic coastline. All the crew members are American but the ship is mostly in international waters. When a ship is in international waters, it has to follow the law of the state whose flag it sails under and each ship can sail under only one flag. This ship does not sail under the U.S. flag. The law of this ship's flag state allows both whipping and food deprivation as a punishment.
On this ship, food deprivation is always used to discipline sailors who disobey orders or who are drunk on duty; as a consequence everyone on this ship has come to think that food deprivation is an appropriate punishment. Whipping however is never used to discipline sailors and no one on this ship thinks whipping is an appropriate punishment. One night, while the ship is in international waters, Mr. Johnson finds a sailor drunk at a time when the sailor should have been on watch. After the sailor sobers up, Mr. Johnson punishes the sailor by giving him 5 lashes with a whip. This does not go against the law of the flag state.

Question: Is it morally permissible for Mr. Johnson/Mr. Williams to whip the sailor?
Yes, it is morally permissible
No, it is morally wrong (whether it is right or wrong in other ways or not)

(2) Mr. Williams is an officer … (same as above)
On this ship, whipping is always used to discipline sailors who disobey orders or who are drunk on duty; as a consequence everyone on this ship has come to think that whipping is an appropriate punishment. Food deprivation however is never used to discipline sailors and no one on this ship thinks food deprivation is an appropriate punishment. One night, while the ship is in international waters, Mr. Williams finds a sailor drunk at a time when the sailor should have been on watch. After the sailor sobers up, Mr. Williams punishes the sailor by giving him 5 lashes with a whip. This does not go against the law of the flag state.

Question: Is it morally permissible for Mr. Williams to whip the sailor?
Yes, it is morally permissible
No, it is morally wrong (whether it is right or wrong in other ways or not)

---

scope of this paper. The design and results reported here relate only to the version of the paired scenarios of their study that, according to them, contain the appropriate way of phrasing the generalizability probe.

Again, the question of the first scenario fulfills the same role of the manipulation-check probe of the standard moral/conventional task while the question of the second scenario is related to the generalizability probe. Quintelier et al.'s probes in fact differed from the ones normally used in the moral/conventional task in two ways. First, they had a third response option, which we did not include above, for each of their questions: "Yes, it is morally permissible but it is wrong for reasons that have nothing to do with morality (e.g., it might be unlawful)." But most participants who chose this option were eliminated from the analysis and the remaining ones were pooled either with the first or second response options described above (see Quintelier et al., 2012). Therefore, this does not have any bearing on our discussion. Second, as one can see above, they introduced "morally" into the questions and response options. We explicate their motivation for doing this in the next subsection.

Again, No-No answers are evidence for our perspective whereas No-Yes answers are evidence against it. The results of Quintelier et al.'s study are presented in the last row of Table 1. Focusing on the difference between the No-Yes and No-Yes' percentages (16% vs. 24%), Quintelier et al. (2012, p. 194) suggest that it is inconsequential to remove participants exhibiting the Yes-Yes response pattern from the analysis, as we prescribe, since this removal does not really make a difference in the results. However, this obfuscates the fact that the removal of these participants makes a big difference in terms of the difference between the No-No and No-No' percentages (51% vs. 76%). Moreover, our prescription to remove participants exhibiting the Yes-Yes response pattern is completely independent of whether this will make any important difference in the results of a particular study, since it is based on the fact that Yes-Yes answers in no way can constitute evidence to test a hypothesis about the folk understanding of harmful *transgressions* qua moral transgressions. Consequently, when the presumed piracy-threat

confound was eliminated in the new study, the great majority of participants (76%) evinced the No-No response pattern. Accordingly, the prima facie evidence coming from the above studies is largely supportive of our perspective.

The descriptive-reading confound. In the studies discussed above, there is still a sizable minority evincing the No-Yes response pattern, which seems to count against our perspective. We would like to show now that this minority is much less significant than one might suppose because (i) the manipulation-check and authority-contingency/generalizability questions utilized in these studies have two different readings—an evaluative reading and a descriptive one; (ii) if participants answer the questions based on a descriptive reading, their answers do not constitute valid evidence; (iii) many No-Yes answers in these studies were based on a descriptive reading.

The questions that constitute both the manipulation-check and authority-contingency/generalizability probes utilized in the studies discussed above have the following general form:

(0) Is it OK/permissible for Mr. X to perform action A?

The intended meaning of the question asks participants to make an *evaluative judgment*. If the question is understood properly, when a participant answers Yes, she is saying that, in performing A, X did not do something wrong, while, when she answers No, she is saying that, in performing A, X did something wrong. In both cases, the participant is evaluating A with her judgment. Let us represent this evaluative understanding of the question by (1).

(1) *Is it OK that Mr. X performs A?*

There is an inappropriate reading of (0) though, which involves simply asking participants to make a *non-evaluative description*. This descriptive understanding of the question is represented by (2).

(2) *According to Y, is it OK that Mr. X performs A?*

'Y' may refer to persons other than the participant of the task, including Mr. X, or to more abstract entities such as tradition, culture, or the legal system. If (0) is understood as (2), when a participant answers Yes, she is saying that, according to Y, in performing A, X did not do something wrong, while when a participant answers No, she is saying that, according to Y, in performing A, X did something wrong. In both cases, the participant herself is not making an evaluative judgment, rather, she is just describing Y's evaluative judgment of A. Thus, if the participant interprets (0) as (2), their Yes or No answer does not constitute valid evidence.

In relation to the above-paired scenarios, we claim both that a non-negligible amount of participants (especially in the whipping-generality pairs) had a descriptive reading of the questions and that this confound is mainly a problem in relation to the No-Yes response pattern. This confound is mainly a problem in relation to the No-Yes pattern due to the asymmetry between the contexts of the first and second scenarios of each paired scenarios reproduced above: in each first scenario, the whipping of the sailor goes against what everyone on the ship accepts as suitable punishment for being drunk on duty; in each second scenario, whipping is endorsed by everyone, or by a relevant authority, as a suitable form of punishment. A descriptive reading of the question therefore can only result in a No-Yes response pattern, as the act of

whipping the drunken sailor is descriptively prohibited in the first scenario and descriptively not prohibited in the second. Participants evincing the No-Yes response pattern via a descriptive reading should have been eliminated from the analysis since their answers do not constitute valid evidence. For this reason, we claim that the percentages of No-Yes answers, which are already relatively low, overestimate the amount of evidence contradicting our perspective, though it is impossible to say exactly how many participants adopted a descriptive reading of the question.

To illustrate the point, consider the justifications of two participants of our study described above (Sousa et al., 2009) evincing a No-Yes response pattern:

Whipping-generality (Participant 1)

Answer to manipulation-check probe: Not-OK
Justification: "I believe receiving lashes for insubordination is too extreme"

Answer to the generalizability probe: OK
Justification: "Such was the standard of the time and what was expected from the sailors and the officers. However it is still wrong."

Whipping-generality (Participant 2)

Answer to manipulation-check probe: Not-OK
Justification: "This is considered to be abusing the crew."

Answer to the generalizability probe: OK
Justification: "It was an acceptable method of keeping Sailor's in line."

The OK answer of the first participant to the generalizability probe was driven by a descriptive reading, since the last remark, "However it is still wrong," indicates the evaluation of the participant—i.e., Not-OK (No). This shows that in reality this participant had a No-No evaluative response pattern, which increases the support for our perspective. The answers of the second participant seem to indicate a descriptive reading of both questions, but one cannot be

absolutely certain whether this was the case. Given this indeterminacy, a reasonable measure would be to eliminate this participant from the analysis, resulting in even less evidence against our perspective.

Acknowledging the relevance of the descriptive-reading confound, Quintelier et al. argue that this was not a problem for their design and results. First (2012, p. 186), they claim that their introduction of "morally" into the questions and response options was meant to lead participants to answer evaluatively. Second (2012, p. 192-193), they attempt to confirm directly that this confound did not affect their results by offering the following justification from a participant who gave a No [Not-OK] answer to the question of the second scenario (i.e., to the generalizability probe):

> "(…) in my mind what is morally right and what is lawfully right are not the same.''

In relation to their first point, there is nothing about the *ordinary meanings* of the words "morally" or "moral" that would lead participants to answer evaluatively. One could easily interpret "morally permissible" as referring simply to what everyone in the ship thinks is or is not acceptable. For example, in the context of the first scenario, a participant may have answered that whipping is "morally wrong" as a description of the normative point of view of the ship, that is, as a description of the fact that on the ship everyone thinks that whipping is an unacceptable form of punishment.

Moreover, in relation to their second point, the usage of the above justification to support their claim shows that Quintelier et al. did not fully understand the problem we have raised and, for this reason, did not focus on the relevant justifications. Since in the context of the second scenario everyone in the ship thinks that whipping is acceptable, we acknowledge that it is quite

plausible to suppose that the above justification indicates an evaluative reading of the question and an evaluative Not-OK answer—that is, that the participant aligns her own evaluation with what she thinks is the moral point of view, *despite* what everyone in the ship thinks (and against the legal point of view too). However, this actually gives support to our claims. This participant would have evinced a No-No response pattern, which is prima facie evidence for our perspective, and confirms that the descriptive-reading problem is not a problem in relation to the No-No response pattern, as we argued above. To support their claim, Quintelier et al. should have provided justifications showing that Yes-answers of the No-Yes response pattern were evaluative. In short, instead of providing evidence for their claim, the quoted justification offered by Quintelier et al. may actually provide evidence for our position.[6]

Thorough evidence. Finally, we claim that No-No and No-Yes response patterns cannot alone test our perspective since prima facie evidence is incomplete evidence—i.e., our perspective cannot be tested independently of evidence concerning the types of inferences guiding participants' Yes or No answers. One way of providing such evidence is to include an analysis of participants' justifications for their answers.

[6] Of course, the descriptive-reading confound constitutes a problem that may compromise the validity and reliability of the moral/conventional task more generally. However, this confound is particularly problematic to the drunken-sailor scenarios discussed here because, in their design (in contraposition to the more traditional design of the moral/conventional task), there is an explicit emphasis on an asymmetry between the institutionalized contexts of the first and second scenarios, which makes the descriptive reading of the questions more salient. Moreover, we would argue that while this problem affects the No-Yes pattern of response related to "harmful" scenarios, it does not affect as much the No-Yes pattern related to "conventional" scenarios, a point that we do not have space to develop here.

The inclusion of an analysis of justifications is particularly important in relation to cases of complex harm, in which a variety of criteria may influence participants' response patterns. Even response patterns that if taken at face value indicate evidence against our perspective may not end up being so when the criteria behind the answers are explicated. Kelly et al. (2007) did not include justification probes in their design, so their studies cannot provide thorough evidence to test our perspective. Quintelier et al. (2012) did include justification probes but their aim was just to check for harm, justice, and rights confounds and their analysis of justifications was not detailed enough to test our hypothesis, given that their coding scheme included only a broad opposition between harm/justice/rights justifications and contextual justifications. On the other hand, we (Sousa et al., 2009) utilized a novel and quite detailed coding scheme to analyze the justifications of Yes or No answers and the results of our analysis clearly support our perspective, as we shall now illustrate.

Our hypothesis says that a harmful transgression is considered to be independent of authority and of social consensus (i.e., considered to be a moral transgression) if it is conceptualized as involving basic-rights violation/injustice. Plain evidence confirming our hypothesis consists of No-No cases in which both No answers are justified in terms of basic-rights violation or injustice. The alternative hypothesis would say that even if a harmful transgression is conceptualized as involving basic-rights violation/injustice, the normative force of an authority or social consensus, independent of the normative content promoted by the authority or social consensus, may cancel the wrongness of the harmful action. Plain evidence confirming the alternative hypothesis (or disconfirming our hypothesis) consists of No-Yes cases in which the No answer is justified in terms of basic-rights violation or injustice and the Yes

answer is justified in terms of an authority or social consensus having an endorsing, normative position towards the act.

The data on justifications concerning our whipping scenarios showed that, of the 54 participants evincing the No-No response pattern, 31 constituted plain evidence for our hypothesis. Moreover, of the remaining participants, 11 justified one of their No answers in terms of basic-rights violation or injustice, which is still important, though more modest, evidence: these participants often justified the other No answer simply by saying that the action was not OK, which might be an economical and implicit way of communicating the same justification of the other No answer (i.e., basic-rights violation or injustice).

On the other hand, of the six participants evincing the No-Yes response pattern, four justified their No answer in terms of basic-rights violation or injustice. Of these participants, two were the participants exhibiting the descriptive-reading confound we discussed earlier. The other two offered the following rationales:

> *Whipping-authority Pair*
>
> Answer to the manipulation-check question: Not-OK
> Justification: "No one has the right to punish another person by inflicting pain…"
>
> Answer to the authority-contingency question: OK
> Justification: "As long as the sailor understood the circumstances in which it may be possible where he would be whipped. He should have a good understanding of the rules and regulations and so should know that not adhering to such would result in punishment."
>
> *Whipping-generality Pair*
>
> Answer to manipulation-check question: Not-OK
> Justification: "Modern times uses more effective means of punishment. Whipping is not expected nor deserved in this situation"
>
> Answer to the generalizability question: OK

> Justification: "Because it was a generally accepted means of punishment that was most likely understood by the sailor"

Three different factors are apparent in these justifications: a utilitarian concern about the efficacy of punishment; a retributive-justice concern about the appropriate type of punishment; a procedural concern about whether the punishment rules are explicit and known. In our interpretation, these participants took these concerns to favor a No answer to each first scenario, but they prioritized the procedural concern in their Yes answer to each second scenario—because they saw the punishment as procedurally appropriate, they deemed it justified. In other words, their Yes answer was not based upon the normative force of an authority or social consensus in and of itself, but on a distinct criterion—namely, a concern for procedure.[7] For this reason, these participants do not constitute plain evidence against our hypothesis (or in favor of the alternative hypothesis). Further, even a sample with a large percentage of No-Yes answers would not constitute such evidence if these answers were based on this distinct criterion.

One may claim that these participants still constitute a modest type of evidence against our perspective, much like the aforementioned 11 participants who justified only one of their No answers in terms of basic-rights violation or injustice (but, of course, quantitatively speaking these participants could not constitute much in terms of evidence—two out of 60 participants is an insignificant fraction). However, in our view, these participants would constitute such modest evidence only if their justification for the Yes answers had left open the possibility that they were implicitly saying that it is OK because of the normative position of an authority or social

[7] Note that this distinct criterion may be related to justice considerations, as punishment-procedural issues have always been discussed under the heading of procedural justice, a point that we don't have space to discuss here.

consensus towards the act—namely, that the authority or social consensus endorses it. However, instead, their justifications indicate a concern distinct from authority/social consensus.

Finally, more generally, one might question whether the "thorough" evidence we described above constitutes *reliable* evidence—i.e., whether participants' justifications can be trusted to accurately reflect the true inferences guiding their permissibility judgments, rather than to simply be post hoc rationalizations of their preceding judgment (see Haidt, 2001). We believe there is good reason to think the justifications in our study represent reliable evidence, in that they did not evince the hallmark symptoms of unreliability: participants did not show difficulties in articulating the reasons for their judgments, there were no apparent contradictions between justifications and patterns of judgments, and, when participants gave multiple justifications, there was an internal coherence that does not suggest confabulation. This provides some assurance that their justifications were more than post hoc rationalizations. For this reason, we claim that the thorough evidence reviewed above constitutes reliable evidence *and* evidence strongly in favor of our deflationary account.

Two further issues

We would like to discuss now two further issues: one concerning the folk understanding of harmful transgressions in particular, and another concerning the folk understanding of moral transgressions in general. As a consequence, we shall envisage two lines of research that might lead to amendments to our perspective.

Harmful transgressions. So far, we have claimed that if a harmful action is seen as involving basic-rights violation/injustice, it is understood to be a transgression, but we have not claimed that a harmful action is understood as a transgression *only if* it is seen as involving basic-rights violation/injustice—i.e., we hypothesized that, according to folk understanding, $\forall x ((Hx$

& *I*x) → (*T*x)), but not that ∀x ((*Hx* & *T*x) → (*I*x)). Here, we would like to discuss whether this further claim might be warranted.

To illustrate, let us pose the following empirical question: are there any cases of harmful actions that people conceptualize as a transgression that they do not conceptualize as involving some basic-rights violation/injustice? Although there are clear ordinary cases of transgressions involving basic-rights violation/injustice without involving harm (e.g., cheating on someone without the person being aware of it), as far as we can tell, there are not any clear cases of harmful transgressions that do not involve some basic-rights violation/injustice. Perhaps committing suicide could constitute such a case. But it is plausible to suppose that when people think that there is an obligation to not commit suicide and hence that committing suicide would be a transgression, they think that suicide would involve a violation of the basic-rights of others in that it would involve the selfish, unjust causation of pain to significant others, or that the victim would be violating their own self-interest (as if the victim were being unfair to themselves) by terminating any opportunity for happiness in the future. Or perhaps the purely accidental causation of harm would constitute such a case (e.g., accidentally dropping a knife on someone's foot). But it is unclear whether people take these cases as an instance of transgression/wrongdoing at all.[8]

Thus, we would like to propose, as an empirically-testable hypothesis, that harmful actions are seen as transgressions *only if* they are seen as involving basic-rights violation/injustice. In other words, it might be the case that, according to folk understanding,

[8] This involves the much more complicated issue about how people understand excuses and the relationship between culpability and wrongdoing, which we do not have space to discuss here.

harmful transgressions are those harmful actions involving basic-rights violation/injustice—i.e., $\forall x ((Hx \ \& \ Tx) \leftrightarrow (Hx \ \& \ Ix))$.

Moral transgressions. So far, we have not claimed that moral transgressions are equivalent to injustice transgressions—i.e., we hypothesized that, according to folk understanding, $\forall x \ (Ix \rightarrow MTx)$, but not that $\forall x \ (Ix \leftrightarrow MTx)$. As we indicated in the introduction, some researches debating with the Turiel tradition have argued against this equivalence, for people seem to categorize transgressions other than basic-rights violation/injustice ones as moral transgressions. Here, we would like discuss whether this equivalence might be warranted.

When researchers claim that people categorize transgressions other than those involving basic-rights violation/injustice as moral transgressions, they are claiming that people can have the same type of strong normative conviction that specifies moral norms in relation to a variety of normative contents—i.e., not only basic-rights violation/injustice transgressions but also some transgressions not involving basic-rights violation/injustice are seen as unquestionably wrong.

There are different ways of characterizing the strong normative conviction that typifies moral norms (e.g., Goodwin & Darley, 2008; Sripada & Stich, 2006; Skitka, Bauman, & Sargis, 2005; Tetlock, 2003; Turiel, 1983). Following the Turiel tradition, we characterized it in terms of the components authority independence and generality. Furthermore, we hypothesized a close conceptual relation between a specific type of normative content and this strong normative conviction in that transgressions deemed to involve basic-rights violation/injustice as their content imply wrongness that is authority independent and general in scope. We would like to hypothesize now that this type of relationship between normative content and strong normative conviction qua authority independence and generality found in relation to basic-rights violation/injustice transgressions is significantly distinct in psychological terms from the

relationship between normative content and strong normative conviction found in transgressions not involving basic-rights violation/injustice. We do not want to deny that people can have a strong normative conviction concerning the latter transgressions; we just want to question whether this strong normative conviction involves authority independence and generality.

Let us illustrate what we have in mind by comparing actions involving basic-rights violation/injustice and actions involving indecency-disgust, such as someone privately masturbating with a dead chicken or having sex with the body of their dead spouse, which have been claimed to be understood as non-injustice moral transgressions (see Haidt et al., 1993; Piazza, Russell, & Sousa, 2013). Because the concept of basic-rights violation/injustice implies transgression, there should be very little variability on whether people conceptualize actions involving injustice as transgressions; by contrast, because disgusting actions do not imply transgression, there should be significant individual and cultural variability on whether these actions are conceptualized as transgressions, and indeed this seems to be the case (e.g., see Piazza & Sousa, in press, Study 3). Because the concept of basic-rights violation/injustice implies authority independence and generality, people should consider transgressions involving basic-rights violation/injustice to be authority independent and general in scope regardless of the level of injustice involved; by contrast, when a disgusting action is conceptualized as a transgression, there will be important individual and cultural variability on whether people have a strong normative conviction about the transgression depending on the level of disgust elicited (and on people's dispositional susceptibility to disgust). Accordingly, the strong normative conviction related to disgusting transgressions should be of a different psychological nature—in particular, it might not involve the components authority independence and generality, but depend more on a person's emotional sensitivities and their ability to regulate these sensitivities.

It may be the case that disgusting transgressions are considered general in scope, as Haidt and colleagues have argued (Haidt et al., 1993). However, take the way generalizability was operationalized in the context of their research:

> Suppose you learn about two different foreign countries. In country A, people [do that act] very often, and in country B, they never [do that act]. Are both of these customs OK, or is one of them bad or wrong.

An answer that both customs are OK would indicate non-generality; an answer that one of them is bad or wrong (presumably the custom where the disgusting act is practiced) would indicate generality. Because the probe was phrased in terms of a custom being "bad or wrong", and because badness does not necessarily imply wrongness, it is unclear whether participants who chose the latter option evinced the component generality. It is plausible that they thought that the custom was simply *suberogatory*—i.e., discretionary, but not recommended.[9] It may also be the case that disgusting transgressions are considered to be authority independent, as Nichols has argued (Nichols, 2002). However, Nichols utilized disgusting actions that were not private, which leaves open the possibility that it was the socially offensive nature of the actions that drove the results (see Royzman, Leeman & Baron, 2009).

Thus, the extent to which the strong normative conviction related to some indecency-disgust transgressions is similar to the strong normative conviction related to basic-rights

[9] For preliminary evidence that ordinary people parse the domain of discretionary actions into suberogatory, neutral, and supererogatory actions, and that this parsing is related to badness/goodness that does not imply wrongness/rightness, see Salomon and Sousa (2010).

violation/injustice transgressions is still an important question that deserves much further empirical research, contrary to the popular assumption in moral psychology that they are equivalent in this respect. It may be that the general notion of strong normative conviction that is normally taken to specify moral norms and moral transgressions subsumes a variety of distinct psychological profiles.

Let us return to the equivalence—i.e., $\forall x\ (Ix \leftrightarrow MTx)$. Whether one uses "moral" as a family-resemblance term covering all strong normative convictions or uses it to refer to a specific subset thereof is not a substantive issue—after all, there is nothing special about the usage of the term "moral". If one uses this term to refer to strong normative convictions in general, it seems indeed that the above equivalence is not warranted. However, if one uses it to refer to a specific type of strong normative conviction characterized in terms of authority independence and generality, as we did, the above equivalence might still be warranted. In other words, leaving aside the unnecessary word "moral", we would like to propose that there might be indeed an equivalence between transgressions seen as involving basic-rights violation/injustice and transgressions whose wrongness is seen as authority independent and general in scope.

Conclusion

In this article, we elaborated and refined the deflationary view on the folk understanding of the relationship between harm and morality we had proposed in our previous work. We also showed that our perspective stands the various criticisms that have been raised against it and is largely supported by the current evidence. Moreover, with our fine-grained approach to data analysis, we demonstrated the type of approach one should pursue to probe our perspective. Finally, we discussed further issues that open new lines of inquiry (or reopen supposedly settled

ones), which in turn might lead to further amendments to our perspective. We would like to conclude by completing our deflationary picture.

So far we have claimed that harmful transgressions are understood as moral transgressions if the causation of harm is interpreted as involving basic-rights violation and injustice, but we have not claimed that harmful transgressions are understood as moral transgressions *only if* the causation of harm is interpreted as involving basic-rights violation/injustice—i.e., according to folk understanding, we have claimed that $\forall x ((Hx \,\&\, Ix) \rightarrow (MTx))$, but not that $\forall x ((MTx \,\&\, Hx) \rightarrow (Ix))$. We did not make this additional claim because the current evidence does not speak to it. However, one may raise the question of whether we would take our perspective to be fully deflationary in that it would incorporate this claim.

If one of the two hypotheses we raised in the previous section ("Two further issues") is true, the additional claim follows suit. If, according to folk understanding, harmful transgressions are those harmful actions involving basic-rights violation/injustice, then harmful transgressions are understood as moral transgressions only if the causation of harm is interpreted as involving basic-rights violation/injustice—i.e., if $\forall x ((Hx \,\&\, Tx) \leftrightarrow (Hx \,\&\, Ix))$, then not only $\forall x ((Hx \,\&\, Ix) \rightarrow (MTx))$, but also $\forall x ((MTx \,\&\, Hx) \rightarrow (Ix))$. If, according to folk understanding, transgressions that are authority independent and general in scope are those transgressions involving basic-rights violation/injustice, then harmful transgressions are understood as moral transgressions only if the causation of harm is interpreted as involving basic-rights violation/injustice—i.e., if $\forall x (Ix \leftrightarrow MTx)$, then not only $\forall x ((Hx \,\&\, Ix) \rightarrow (MTx))$, but also $\forall x ((MTx \,\&\, Hx) \rightarrow (Ix))$. However, if neither hypothesis is true, there is still the possibility that some harmful transgressions not involving basic-rights violation/injustice are understood as authority independent and general in scope.

We do not have clear evidence in favor of the two hypotheses we raised in the previous section, nor do we have such evidence regarding the aforementioned possibility that is opened up if these two hypotheses are false. Nonetheless, we would like to close by saying that, even if these two hypotheses are false, we anticipate that future research will bear out that harmful transgressions are understood as moral transgressions *if and only if* the causation of harm is interpreted as involving basic-rights violation/injustice. In short, we anticipate that the fully deflationary view will be the correct one.

References

Baumard, N., André, J.B., & Sperber, D. (2013). A mutualistic approach to morality. *Behavioral and Brain Sciences* (target article), *36*, 59-122.

Beller, S. (2008a) Deontic norms, deontic reasoning, and deontic conditionals. *Thinking and Reasoning*, *14*, 305-341.

Beller, S. (2008b). Deontic reasoning squared. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 2103-2108). Austin, TX: Cognitive Science Society.

Bender, A., & Beller, S. (2003). Polynesian tapu in the 'deontic square': A cognitive concept, its linguistic expression and cultural context. In R. Alterman & D. Kirsh (Eds.), *Proceedings of the 25th Annual Conference of the Cognitive Science Society* (pp. 133–138). Mahwah, NJ: Lawrence Erlbaum Associates Inc.

Bucciarelli, M., & Johnson-Laird, P. N. (2005). Naive deontics: A theory of meaning, representation, and reasoning. *Cognitive Psychology*, *50*, 159–193.

Cushman, F. A., Gray, K., Gaffey, A., & Mendes, W. B. (2012). Simulating murder: The

aversion to harmful actions. *Emotion, 12,* 2-7.

Fraser, B. (2012). The nature of moral judgments and the extent of the moral domain. *Philosophical Explorations: An International Journal for the Philosophy of Mind and Action*, *15*, 1–16.

Goodwin, G.P., and J.M. Darley. 2008. The psychology of meta-ethics: exploring objectivism. *Cognition*, *106*, 1339–1366.

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

Haidt, J., & Joseph, C. (2007). The moral mind: how five sets of innate intuitions guide the development of any culture-specific virtures, and perhaps even modules. In P. Carruthers, S. Laurence, & S. Stich (Eds.) *The innate mind* (Vol. 3, pp. 367-391). New York: Oxford University Press.

Haidt, J., Koller, S., & Dias, M. (1993). Affect, culture and morality, or is it wrong to eat your dog? *Journal of Personality and Social Psychology*, *65*, 613-628.

Helwig, C., Hildebrandt, C., & Turiel, E. (1995). Children's Judgments about Psychological Harm in Social Context. *Child Development*, *66*, 1680.

Jackendoff, R. (1999). The natural logic of rights and obligations. In R. Jackendoff, P. Bloom, & K. Wynn (Eds.), *Language, Logic, and Concepts – essays in memory of John Macnamara* (pp. 67-95). Cambridge: The MIT Press.

Kelly, D., Stich, S., Haley, K., Eng, S., & Fessler, D. (2007). Harm, affect, and the moral/conventional distinction. *Mind and Language*, *22*, 117-131.

Lycan, W. (1986). Tacit belief. In R.J. Bogdan (Ed.), *Belief: Form, content, and function* (pp. 61–82). Oxford: Clarendon.

Machery, E., & Mallon, M. (2010). The evolution of morality. In J. M. Doris & The Moral Psychology Research Group (Eds.), *The moral psychology handbook* (pp. 3–46). Oxford, England: Oxford University Press.

Nichols, S. (2002). Norms with feeling: Towards a psychological account of moral judgment. *Cognition*, *84*, 221–236.

Nichols, S. (2004). *Sentimental rules: On the natural foundations of moral judgment*. New York: Oxford University Press.

Nucci, L. (2001). *Education in the moral domain*. Cambridge: Cambridge University Press.

Nucci, L., & Turiel, E. (1978). Social interactions and the development of social concepts in preschool children. *Child Development*, *49*, 400-407.

Piazza, J.**,** Russell, P. S., & Sousa, P. (2013). Moral emotions and the envisioning of mitigating circumstances for wrongdoing. *Cognition & Emotion, 27,* 707-722.

Piazza, J., & Sousa, P. (in press). Religiosity, political orientation, and consequentialist moral thinking. *Social Psychological and Personality Science*.

Piazza, J., Sousa, P., & Holbrook, C. (2013). Authority dependence and judgments of utilitarian harm. *Cognition*, *128*, 261-270.

Quintelier, K., Fessler, D., & De Smet, D. (2012). The case of the drunken sailor: On the generalizable wrongness of harmful transgressions. *Thinking and Reasoning*, *18*, 183-195.

Royzman, E., Leeman, R. F., & Baron, J. (2009). Unsentimental ethics: Towards a content-specific account of the moral–conventional distinction. *Cognition*, *112*, 159–174.

Salomon, E., & Sousa, P. (2010). *Beyond wrongdoing: How the folk parse the moral domain.* Poster presented at the meetings for the Society for Philosophy and Psychology, Montreal, Canada.

Samuels, R., & Stich, S. (2004). Rationality and psychology. In A. Mele & P. Rawling (Eds.), *The Oxford handbook of rationality* (pp. 279–300). Oxford: Oxford University Press.

Shweder, R. A., Mahapatra, M., & Miller, J. (1990). Culture and moral development. In J. Stigler, R. Shweder, and G. Herdt (Eds.), *Cultural Psychology – essays on comparative human development* (pp. 130-204). Cambrige: Cambridge University Press.

Singer, P. (2011). *The expanding circle: Ethics, evolution and moral progress.* Princeton: Princeton University Press.

Skitka, L. J., Bauman, C. W., & Sargis, E. G. (2005). Moral conviction: Another contributor to attitude strength or something more? *Journal of Personality and Social Psychology, 88,* 895 – 917.

Smetana, J. (1981). Preschool children's conceptions of moral and social rules. *Child Development*, *52*, 1333 -1336.

Smetana, J. (1993). Understanding of social rules. In M. Bennet (Ed.), *The Development of Social Cognition: The Child as Psychologist*. New York: Guilford Press.

Smetana, J., & Braeges, J. (1990). The development of toddlers' moral and conventional judgements. *Merrill-Palmer Quarterly, 36*, 329-346.

Sousa, P., Holbrook, C., & Piazza, J. (2009). The morality of harm. *Cognition*, *113*, 80-92.

Sousa, P. (2009a). On testing the moral law. *Mind & Language*, 24, 209-234.

Sousa, P. (2009b). A cognitive approach to moral responsibility—the Case of a Failed Attempt to Kill. *Journal of Cognition and Culture*, *9*, 171-194.

Sperber, D., & Baumard, N. (2012) Morality and reputation in an evolutionary perspective, *Mind and language, 27*, 495-518.

Sripada, C. S., & Stich, S. (2006). A framework for the psychology of norms. In P. Carruthers, S. Laurence, & S. Stich (Eds.), *The innate mind* (Vol. 2, pp. 280–301). New York: Oxford University Press.

Stich, S., Fessler, D., & Kelly, D. (2009). On the morality of harm: A response to Sousa, Holbrook and Piazza. *Cognition*, *113*, 93-97.

Stich, S., Doris, J. & Roedder, E. (2010). Altruism. In Moral Psychology Research Group (Ed.), *The Oxford Handbook of Moral Psychology* (pp. 147-20). Oxford: Oxford University Press.

Tetlock, P. E. (2003). Thinking the unthinkable: Sacred values and taboo cognitions. *Trends in Cognitive Sciences*, *7*, 320–324.

Tisak, M. (1995). Domains of social reasoning and beyond. In R. Vasta (Ed.), *Annals of Child Development* (Vol. 11, pp. 95–130). London: Jessica Kingsley.

Tisak, M., & Turiel, E. (1984). Children's conceptions of moral and prudential rules. *Child Development*, *55*, 1030-1039.

Turiel, E. (1983). *The development of social knowledge*. Cambridge: Cambridge University Press.

Turiel, E. (2002). *The culture of morality*. Cambridge: Cambridge University Press.

Turiel, E., Hildebrandt, C., & Wainryb, C. (1991). Judging social issues: Difficulties, inconsistencies and consistencies. *Monographs for the Society of Research in Child Development, 56* (Serial no. 224).

Wainryb, C. (1991). Understanding differences in moral judgments: The role of

informational assumptions. *Child Development*, 62, 840-851.

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

Weston, D. R., & Turiel, E. (1980). Act-rule relations: Children's concepts of social rules. *Developmental Psychology, 16,* 417-424.

Yau, J., & Smetana J. 2003: Conceptions of moral, social -conventional, and personal events among Chinese preschoolers in Hong Kong. *Child Development*, *74*, 647-658.