

On Testing the ‘Moral Law’

PAULO SOUSA

Abstract: In a previous article in this journal, Daniel Kelly, Stephen Stich, Kevin Haley, Serena Eng and Daniel Fessler report data that, according to them, foster scepticism about an association between harm and morality existent in the Turiel tradition (Kelly *et al.*, 2007). This article challenges their interpretation of the data. It does so by explicating some methodological problems in the Turiel tradition that Kelly *et al.* themselves in a way inherit and by drawing on new evidence coming from a partial replication of their research.

1. Introduction

Over the last three decades, psychologists Elliot Turiel, Larry Nucci, Judith Smetana and colleagues have developed an influential paradigm on the structure of the concept of moral wrongdoing (e.g. Nucci, 2001; Smetana, 1993; Tisak, 1995; Turiel, 1983), a paradigm whose basic tenets have been adopted by other influential authors such as R. J. R. Blair and Shaun Nichols (e.g. Blair, 1995; Nichols, 2004). In a previous article in this journal (Kelly *et al.*, 2007; see also Kelly and Stich, in press), Daniel Kelly, Stephen Stich, Kevin Haley, Serena Eng and Daniel Fessler raise scepticism concerning an association between harm and morality present in this paradigm, an association I shall call the ‘moral law’.¹ One fundamental methodology utilized by the Turiel tradition is the moral/conventional task, a methodology that in a way the harm sceptics follow in their research designed to test the moral law.

The goal of this article is to question the claims of the harm sceptics with the help of new data coming from a partial replication of their research (for a detailed

Thanks to the AHRC *Culture and the Mind* project, under the scope of which I became aware of the issues discussed in this article, and to Nicholas Baumard, George Botterill, Fabrice Clement, Graham MacDonald, Daniel Kelly, Nicola Knight, Robert N. McCauley, Shaun Nichols, Dan Sperber and Monica Whitty for their helpful comments. Special thanks to Colin Holbrook and Jared Piazza for their invaluable feedback on all aspects of the article.

Address for correspondence: Institute of Cognition and Culture, Queen’s University, Belfast, 2-4 Fitzwilliam Street, Belfast, BT71NN, UK.

Email: p.sousa@qub.ac.uk

¹ In this article, whenever I use the expression ‘moral law,’ I use the word ‘law’ in a descriptive sense akin to usages such as the laws of human behaviour or the laws of nature, never in its prescriptive sense. I use the expression ‘harm sceptics’ to refer to Kelley *et al.*, and the expression ‘Turiel tradition’ in a broad sense that includes not only the work of Turiel, Nucci, Smetana and colleagues but also the work of authors like Blair and Nichols who would accept that their paradigm is on the right track at least in delimiting what is *core* morality.

empirical report of the replication, see Sousa *et al.*, forthcoming). In section two, I sketch the Turiel tradition's paradigm and the harm sceptics' scepticism. Next, I discuss the research of the harm sceptics by raising my meta-scepticism, which is based on problems concerning data analysis (section three), task validity and reliability (section four), and the proper scope of the moral law (section five). In the conclusion, I elucidate the broader significance of my meta-scepticism.

2. The Dispute

In this section, I first delineate the moral domain as understood by the Turiel tradition and one fundamental methodology utilized to test its psychological reality. Then, I sketch the harm sceptics' doubts about the association between harm and morality coming from this tradition.

Despite the fact that a general distinction between moral and non-moral domains does not by itself imply that the non-moral domain is homogeneous (Fiddick, 2004; Tisak and Turiel, 1984), the Turiel tradition has delimited the category *moral* mainly in opposition to the more specific category *conventional*. Lay intuition seems to differentiate transgressions such as *killing* and *stealing*, which, with the norms forbidding them, would be part of the 'moral' category, from transgressions such as *drinking soup out of a bowl* and *wearing inappropriate clothing*, which, with the norms forbidding them, would be part of the 'conventional' category. Drawing originally on philosophical analyses of both conceptual sides of this intuitive divide, the Turiel tradition has hypothesized that the differential values of various conceptual dimensions delimit the two categories (e.g. Blair, 1995; Nichols, 2004; Nucci, 2001; Smetana, 1993; Turiel, 1983). In this article, I deal only with the dimensions that speak to the controversy with the harm sceptics and to the problems I shall explicate in the following sections. These include:

TYPE OF ACTION

- Moral transgressions involve a victim being harmed and being subjected to injustice and/or rights violation.
- Conventional transgressions do not involve such a victim.

'PERMISSIBILITY'

- Moral transgressions are less 'permissible' than conventional transgressions.

SERIOUSNESS

- Moral transgressions are worse than conventional transgressions.

AUTHORITY CONTINGENCY

- Moral transgressions (and the norms forbidding them) are independent of authority—their wrongness is not cancellable by the decision of any authority.
- Conventional transgressions (and the norms forbidding them) depend on authority—their wrongness is cancellable by the decision of a legitimate authority.

GENERALITY

- Moral transgressions (and the norms forbidding them) are general in scope—their wrongness extends to different places and times.
- Conventional transgressions (and the norms forbidding them) are local—their wrongness is local.

It is important to highlight the special status of the first dimension: the differential quality of the two types of transgressive actions, including the broader events they are part of, prompts the differential values of the other dimensions (Turiel, 1983, especially chapters 3 and 4). Let me symbolize this differential quality by using 'H' for harmful action and 'N' for *non*-harmful action.

The moral/conventional task is a methodology utilized to test the psychological reality of the moral/conventional distinction. Hs such as *a child hitting another child* and *a child stealing another child's apple*, and Ns such as *a boy wearing nail polish* and *a child eating lunch with fingers* are described in a random order to each participant in the task. They are described neither as transgressions/wrongdoings nor as moral or conventional transgressions/wrongdoings, but simply as something someone does. For each instance of H or N, a sequence of questions is posed. The questions are supposed to probe the psychological relevance of the conceptual dimensions aforementioned. The standard task goes like this:²

Action Scenario

H (N)

1. 'Permissibility' probe
Is it OK for X to H (N)? YES NO
2. Seriousness probe
How bad is it for X to H (N)? (NOT AT ALL) 0 1 2 3 4 5 (VERY BAD)
3. Justification probe
Why is it bad for X to H (N)?
4. Authority contingency probe
Now, what if an authority says that it is OK to H (N). Would it be OK for X to H (N)? YES NO
5. Generality probe
In another place and/or time, is it OK to H (N)? YES NO

² In the specific versions of this standard, sometimes one of these questions is excluded (and others are included), sometimes the justification probe comes after (and is related to) the permissibility probe, sometimes the order of the two last probes is reversed, and sometimes the word 'wrong' or the expression 'all right' substitutes for the word 'OK' in some of the questions; whichever version though, the order of the presentation of the questions is fixed (cf. Blair, 1995; Nichols, 2002, 2004; Smetana, 1981, 1986; Smetana and Braeges, 1990; Smetana *et al.*, 1993). Other designs depart to a lesser or greater extent from this standard (e.g. Weston and Turiel, 1980, which also includes scenarios with actions described as transgressions; Smetana, 1985, where the scenarios do not specify the actions; Nucci, 1981, where a pile-sorting procedure is used; Nucci and Turiel, 1978, where observational techniques involving less structured interviews are used).

The prediction of the researchers of the Turiel tradition is that participants will judge instances of H (e.g. *a child hitting another child*) as very bad, not-OK even when an authority says OK, and not-OK in others places and/or times, whereas they will judge instances of N (e.g. *a boy wearing nail polish*) as not too bad, OK when an authority says it is OK, and OK in other places and/or times.³ Furthermore, participants will tend to justify their judgements on the badness of the former type in terms of harm, injustice and/or rights violation, while they will not justify their judgements on the badness of the latter in such terms, which would confirm that their judgements are related to the differential quality of the two types of actions. Presumably, if participants dissociate their judgements as predicted, this is strong evidence that they make the distinction between morality and convention characterized by the Turiel tradition.

According to the Turiel tradition, in about three decades of research, the general finding is that judgements on Hs and Ns differ systematically and in the predicted direction, with a similar pattern of results being evinced in a diverse range of participants in terms of age (e.g. Smetana, 1981; Smetana and Braeges, 1990), of religious or more general cultural background (e.g. Nucci *et al.*, 1983; Nucci and Turiel, 1993; Yau and Smetana, 2003), and of psychological abnormality (e.g. Blair, 1996; Nucci and Herman, 1982; Smetana *et al.*, 1984), though children with psychopathic tendencies and adult psychopaths seem to be an exception to the general pattern (e.g. Blair, 1995, 1997). Accordingly, the Turiel tradition claims that human beings in general possess the concepts of morality and convention as characterized.

The harm sceptics refer to the above two opposite patterns of response as the moral and conventional signatures (Kelly *et al.*, 2007, p. 19). According to them, given the cross-cultural recurrence and the early ontogenetic emergence of the task's results, these two signatures seem to constitute 'nomological clusters'—there seems to be a strong tendency for the elements of each of the signatures to co-occur. Now, from the fact that transgressions H evoke the moral signature, and transgressions N evoke the conventional signature, two probabilistic laws seem to hold: respectively, *if transgression H, then moral signature; if transgression N, then conventional signature*. Finally, it seems that behind these laws are two functionally distinct mechanisms of the mind that correspond to the theoretical entities of the model of the Turiel tradition—the conceptual systems of morality and convention.

However, the harm sceptics deny that what looks like to be the case is actually the case (Kelly *et al.*, 2007, pp. 120–121). They claim that the influential work of Jonathan Haidt and others has already provided enough evidence showing that the two signatures are not nomological clusters, and that the conventional law *if*

³ I don't mention here the prediction related to the results of the 'permissibility' probe because, within the Turiel tradition, there is some disagreement in this respect (see my discussion in the next section).

transgression N, then conventional signature does not hold (e.g. Haidt *et al.*, 1993; cf. Gabennesch, 1990). The main focus of their own empirical research is on the moral law *if transgression H, then moral signature*. They acknowledge that the current evidence does confirm the moral law,⁴ but they think this is due to the fact that only a very narrow range of instances of Hs has been utilized in the moral/conventional task: '... all of the harmful transgressions studied have been of the "schoolyard" variety, even when the experimental subjects were incarcerated psychopathic murderers (Blair, 1995)!' (Kelly *et al.*, 2007, p. 121). They suspect that, if different instances of transgressions H were included, many results like *transgression H & not moral signature* would occur, thereby disconfirming the moral law to a substantial extent; and they claim that in their research, by pursuing this inclusion, their suspicion is borne out.

The overall purpose of the next three sections is to argue, with the help of a partial replication of the research of the harm sceptics, that the evidence against the moral law coming from their research is not as substantial as they claim—i.e. to raise my meta-scepticism.

3. Data Analysis Problems

In this section, I deal with two problems of data analysis, which I call the permissibility and pooling problems. First, I explicate the permissibility problem. Then, I describe the research of the harm sceptics and the pooling problem. Finally, I show that, when the data of the harm sceptics is reanalyzed in a way that eliminates these two problems, the evidence against the moral law is not as substantial as they claim.

Within the Turiel tradition, most of the time, 'permissibility' is neither taken as a conceptual dimension whose differential values would distinguish morality and convention, nor is its probe intended to inform on such a distinction, as these passages indicate:

Because all transgressions are, by definition, not permissible, it was expected that children in the present study do not distinguish moral and conventional transgressions on this basis ... (Smetana and Braeges, 1990, p. 331).

The moral/conventional task gets its interest primarily because it gives us a glimpse into judgements of wrong. This is reflected by the fact that the items in the moral/conventional task are explicitly transgressions, and the first question in standard moral/conventional tasks checks for the permissibility of the action (Nichols, 2002, p. 224).

⁴ Notice that some important part of the evidence coming from Haidt and others is of the type *transgression N & moral signature*, which would also disconfirm the moral law if one were to characterize it in terms of a biconditional sentence. Thus, to acknowledge that the current evidence confirms the moral law coherently, the moral law has to be characterized in terms of a conditional sentence.

But ‘permissibility’ has been, even if with some caveat, interpreted as a distinguishing dimension and probe, notably by Blair and Nichols, as these passages indicate:

What is striking about this literature is that, from a young age, children distinguish the moral violations from the conventional violations on a number of dimensions. For instance, children tend to think that moral transgressions are generally less permissible and more serious than conventional transgressions (Nichols, 2002, p. 202).

... while all of the transgression situations, whether moral or conventional, are generally judged not permissible, conventional transgressions are more likely to be judged permissible than moral transgressions ... (Blair, 1995, p. 6).

Let me name the first position ‘the mainstream position’ and the second ‘the minority position’.

I think the mainstream position is right. Here is a somewhat explicit argument. The word ‘permissible’ is polysemous: it has a specific sense that refers to what is discretionary and a superordinate sense that includes in its extension both what is discretionary and what is obligatory.⁵ By definition, transgressions/wrongdoings cannot be permissible in any of these two senses, since they constitute the extension of what is forbidden, which is neither discretionary nor obligatory. Therefore, neither moral nor conventional transgressions/wrongdoings can be permissible in any of the above senses, and permissibility cannot be a distinguishing dimension/probe. True, the word ‘permissible’ seems to have an additional sense corresponding to the idea that a transgression does not involve a great amount of (harsh) punitive consequences for the wrongdoer, which may lead one to say that conventional transgressions are more permissible than moral transgressions. That this meaning is not equivalent to the dimension of permissibility is attested by the fact that a probe on punishment, different from (and in addition to) the permissibility probe, is often included in the moral/conventional task.

I think, though, that both mainstream and minority positions run into a methodological problem. To understand the problem, it is important to distinguish the point of view of the participant who takes the task from the point of view of the researcher who devises the task. The researcher selects Ns and Hs that she categorizes as transgressions, expecting that participants will also categorize them as transgressions, since the aim of the task is to test whether participants dissociate two categories of transgressions (conventional versus moral transgressions). Nevertheless, there is no logical or empirical necessity that participants will judge

⁵ The discretionary sense of the word ‘permissible’ (permissible in the superordinate sense but not obligatory) may be considered a case of Gricean generalized implicature—see Grice, 1991 [1968].

that a specific action N or H, classified by the researcher as a transgression, is a transgression. This is not only because in the standard task the actions are described simply as something someone does. Even if the action were described as a *transgression*, there is no such necessity, since participants' judgements do not have to agree with the point of view of the description—they may judge that the action being described is in fact discretionary, or even obligatory. For this reason, the permissibility probe should be treated as a question checking whether the participant in reality categorizes the action as a transgression—i.e. as a manipulation check. In the moral/conventional task, the answers of a participant to the sequence of probes of a specific action-stimulus scenario constitute a sequence of data points. To treat the 'permissibility' probe as a manipulation check in a strict sense means that one would have to eliminate from the final analysis of the results all sequences of data points where participants answered OK to the 'permissibility' probe, because an OK answer to the permissibility probe indicates that the participant does not consider the action as a transgression. Imagine a participant, after having answered OK to the 'permissibility' probe 'Is it OK for X to N (e.g. a child eating lunch with fingers)?', being asked the authority contingency probe 'Now, what if an authority says that it is OK to N. Would it be OK for X to N?'. Besides the fact that the subsequent question is pragmatically awkward given the first answer, there is the more critical point that any answer to it is completely irrelevant given its original motivation, which is to test whether the OK of an authority would make a participant change a judgement of wrongdoing. So, what could be an epistemologically sound reason to not eliminate from the final analysis of results the sequences of data points initiated by an OK answer to the permissibility probe?

However, both mainstream and minority positions *do not* normally treat the results of the 'permissibility' probe as a manipulation check, when they *ought to do so* (e.g. Blair 1995, Nichols 2002, Smetana and Braeges 1990). This is what I call the permissibility problem. This problem is common to the Turiel tradition as a whole, even if, by not in practice considering the permissibility probe as a manipulation check, the mainstream position is not following the methodological implications of its correct understanding of the permissibility dimension/probe, while the minority position is just following the implications of its misunderstanding.

As I shall show, the permissibility problem affects the harm sceptics too, but in a puzzling way, given that they run first into another problem, which I call the pooling problem. I turn now to the research of the harm sceptics intended to test the moral law.

In order to test the moral law *if transgression H, then moral signature*, their research includes a broader range of Hs and focuses on the elements of the moral signature related to the dimensions of generality and authority contingency, but separately. They created pairs of brief scenarios with different Hs. In each pair, a specific H is kept constant and either the dimension of generality or the dimension of authority contingency is manipulated across the two scenarios of the pair. Then, in a study

conducted on the Internet, only one of these brief pairs was presented to each participant, who responded to the two scenarios of the pair one at a time.⁶ The two generality related pairs with their respective questions are:

SLAVERY GENERALITY

- (1) In the United States, slaves were an important part of the economy of the South 200 years ago. American slaves were used mainly to maintain households, and to supply agricultural labor.

Question: Was it OK for Americans to keep slaves? YES NO

- (2) In ancient Greece and Rome, slaves were an important part of the economic and social system. Greek and Roman slaves were used as oarsmen, as soldiers, to maintain households, and to supply agricultural labor.

Question: Was it OK for the ancient Greeks and Romans to keep slaves? YES NO

WHIPPING GENERALITY

- (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.

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

Three of the authority contingency related pairs with their respective questions are:

WHIPPING AUTHORITY

- (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

⁶ Their overall research included nine pairs of scenarios. One of them is unrelated to the moral law. Of the other pairs, three include instances of H quite similar to the ones utilized in the Turiel tradition. In keeping with their own rationale for testing the moral law, I discuss only the remaining five scenarios, which are the most atypical as far as the Turiel tradition is concerned. In each of the scenarios, there was a second question probing the dimension of seriousness, which I do not discuss in this article.

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

PRISONER AUTHORITY

- (1) Sergeant Johnson is interrogating a suspected terrorist who may have information about future terrorist attacks. His commanding officer has ordered him not to use sleep deprivation as a way of getting information. Nonetheless Sergeant Johnson keeps the suspect awake for three days and three nights.

Question: Is it OK for Sergeant Johnson to keep the suspect awake for three days and three nights? YES NO

- (2) Now suppose that before he decided to keep the prisoner awake, Sergeant Johnson's commanding officer had told him that the use of sleep deprivation is an acceptable way of trying to get information when interrogating suspected terrorists, and that Sergeant Johnson could use sleep deprivation whenever he wanted to.

Question: Is it OK for Sergeant Johnson to keep the suspect awake for three days and three nights? YES NO

TRAINING AUTHORITY

- (1) For many years, the military training of elite American commandos included a simulated interrogation by enemy forces in which the trainees were threatened and physically abused. Most people in the military believe that these simulated interrogations were helpful in preparing trainees for situations they might face later in their military careers. Though no one was ever killed or permanently disabled by the physical abuse they received during these simulated interrogations, the trainees often ended up with bruises or injuries that lasted for a week or more. Recently, the Pentagon issued orders prohibiting physical abuse in military training. Sergeant Anderson is a soldier who trains elite American commandos. He knows about the orders prohibiting physical abuse and his immediate superiors have ordered him not to do it. Nonetheless, he regularly threatens and physically abuses trainees during the simulated interrogations that he conducts.

Question: Is it OK for Sergeant Anderson to physically abuse trainees during simulated interrogations? YES NO

- (2) Now suppose that the Pentagon had never issued orders prohibiting physical abuse in military training, and that Sergeant Anderson's superiors had told him that the use of physical abuse was acceptable in simulated interrogations.

Question: Is it OK for Sergeant Anderson to physically abuse trainees during simulated interrogations? YES NO

Although this design departs from the standard moral/conventional task in many respects, each first scenario of these five pairs presents a case of H that presumably participants would judge an instance of transgression/wrongdoing. Hence, these first scenarios and their respective questions have a similar role to the initial scenario description and the permissibility (manipulation-check) probe of the standard task. Now, each second scenario/question of these five pairs has a similar role to one of the subsequent probes of the task. In the authority pairs, since the second scenarios describe an authority officially cancelling the transgressive character of the Hs of the first scenarios, their questions are intended to probe whether participants will judge that the wrongness of the Hs is contingent on that authority. In the generality pairs, since the second scenarios describe the Hs of the first scenario in a very distant spatial and/or temporal context, their questions are intended to probe whether participants will judge that the wrongness of the Hs extends to that context. Thus, in relation to the generality pairs, for example, the harm sceptics say:

According to (C-2a) [the moral law], transgressions involving harm evoke the signature moral pattern, and one component of that pattern is generality: actions judged wrong here and now should also be judged wrong at other times and in other cultures. *So if transgressions involving harm are regarded as wrong if they are committed now (or in recent history) but are not judged to be wrong if they were committed long ago, this poses a direct challenge to (C-2a) [the moral law].* Two of the scenario sets [pairs] were designed to explore whether participants generalize their responses to transgressions of harm norms that are quite different from the schoolyard harm norms and transgressions typically used in moral/conventional task studies (Kelly *et al.*, 2007, p. 123; my emphasis).⁷

The harm sceptics claim that their results give substantial disconfirmation to the moral law. They base their claim on the fact that, in each of the above five pairs of scenarios, there was a statistically significant difference between the percentage of YES answers (i.e. OK answers) to the first scenario and the percentage of YES answers (i.e. OK answers) to the second in the direction they predicted (more YES answers to the second scenarios), as depicted in Table 1.⁸

⁷ Notice that, in the context of the moral law, keeping slaves in the first scenario of the slavery generality pair (200 years ago in the US) is being considered to be under the scope of recent history while whipping in the second scenario of the authority generality pair (300 years ago) is being considered under the scope of long ago.

⁸ In their study, the order of presentation of the scenarios of each pair was counter-balanced. Because this factor is irrelevant to my arguments in this article, I discuss their results with the two orders of presentation pooled (as the harm sceptics themselves discuss their results, by the way).

| Pair of Scenario | (N) | (1) | (2) |
|---------------------|-----|-----|-----|
| Slavery Generality | 188 | 7% | 11% |
| Whipping Generality | 198 | 10% | 51% |
| Whipping Authority | 196 | 6% | 22% |
| Prisoner Authority | 173 | 1% | 15% |
| Training Authority | 150 | 9% | 58% |

Table 1 (N) Total number of participants
 (1) Percentage of who said YES in scenario 1
 (2) Percentage of who said YES in scenario 2

I don't think the way the harm sceptics justify their claim is convincing. In fact, in pooling participants' answers to each scenario of a pair and focusing solely on the pooled difference between the two scenarios of each pair, they do not report the data that could show the extent to which the moral law is disconfirmed in each of the pair of scenarios. Obviously, this pooling of the data could not be considered a case of a legitimate scientific idealization—hence, the pooling problem.

In their design, a participant who completed the two OK questions of a pair of scenarios could show one of the following patterns of response: (A) NO to the first scenario and NO to the second; (B) NO to the first scenario and YES to the second; (C) YES to the first scenario and YES to the second; (D) YES to the first scenario and NO to the second. Now, if the OK question of the first scenario constitutes the permissibility probe, which should function as a manipulation check, participants who evinced the patterns of response (C) and (D) should be eliminated of the analysis for their OK/YES answer indicates that they did not consider the action as a transgression/wrongdoing. In terms of the moral law *if transgression H, then moral signature (not-OK even when an authority says OK and not-OK in others places and/or times)*, patterns (C) and (D) are irrelevant to test the moral law because they indicate that the participant did not judge H as a transgression and therefore that the antecedent of the moral law is not instantiated. Patterns (A) and (B) are relevant for the opposite reason. Pattern (A) is partial evidence confirming the moral law because any of its two versions, *transgression H & Not-OK in other places and/or times* and *transgression H & Not-OK when an authority says OK*, affirms only one of the conjuncts of the moral signature, therefore not the moral signature as a whole. Pattern (B) is evidence disconfirming the moral law because any of its two versions, *transgression H & OK in other places and/or times* and *transgression H & OK when an authority says OK*, negates one of the conjuncts of the moral signature, therefore the moral signature as a whole. Table 2 contains the information of Table 1 plus the results of a replication of the harm sceptics' research, including the missing information to test the moral law (or what might be such information in the case of their research).⁹

⁹ The replication followed exactly their web-based methodology (see Sousa *et al.*, forthcoming), but the seriousness probe of their design was replaced by a question requiring participants to justify their answers to the OK questions.

| Pair of Scenario | (N) | (1) | (2) | (A) | (B) | (C) | (D) | (A') | (B') |
|---------------------|-----|-----|-----|-----|-----|-----|-----|------|------|
| REPLICATION | | | | | | | | | |
| Slavery Generality | 30 | 3% | 3% | 97% | 0% | 3% | 0% | 100% | 0% |
| Whipping Generality | 30 | 7% | 23% | 77% | 16% | 7% | 0% | 82% | 18% |
| Whipping Authority | 33 | 3% | 6% | 94% | 3% | 3% | 0% | 97% | 3% |
| Prisoner Authority | 35 | 6% | 20% | 77% | 17% | 3% | 3% | 82% | 18% |
| Training Authority | 31 | 16% | 64% | 36% | 48% | 16% | 0% | 42% | 58% |
| HS' RESEARCH | | | | | | | | | |
| Slavery Generality | 188 | 7% | 11% | 89% | 4% | 7% | 0% | 95% | 5% |
| Whipping Generality | 198 | 10% | 51% | 49% | 41% | 10% | 0% | 54% | 46% |
| Whipping Authority | 196 | 6% | 22% | 78% | 16% | 6% | 0% | 83% | 17% |
| Prisoner Authority | 173 | 1% | 15% | 85% | 14% | 1% | 0% | 86% | 14% |
| Training Authority | 150 | 9% | 58% | 42% | 49% | 9% | 0% | 46% | 54% |

Table 2 (N) Total Number of participants

(1) Percentage of who said YES in scenario 1

(2) Percentage of who said YES in scenario 2

(A) Percentage of who said NO in scenario 1 and NO in scenario 2

(B) Percentage of who said NO in scenario 1 and YES in scenario 2

(C) Percentage of who said YES in scenario 1 and YES in scenario 2

(D) Percentage of who said YES in scenario 1 and NO in scenario 2

(A') Percentage of (A) if irrelevant participants are eliminated

(B') Percentage of (B) if irrelevant participants are eliminated

The pooled results of the replication, represented in columns (1) and (2), go in the same direction of their results in most pairs of scenarios, but are as misleading as theirs. The important results are on the other columns on the right. Almost no participant showed pattern (D)[YES-NO], as one would obviously expect: in the authority pairs, why, saying YES in the first scenario, would one say NO in the second, where an authority is supporting one's own judgement? In the generality pairs, why, saying YES to whip now (or to keep slaves in the American South), would one say NO to whip 300 years ago (or to keep slaves in Greece/Rome)? Some participants showed pattern (C)[YES-YES] in each of the pairs of scenarios. These two patterns of response are irrelevant to test the moral law, since they do not instantiate its antecedent. If one eliminates the participants of these two patterns from the analysis in each pair of scenarios, and, with the new total number, recalculates the percentages of patterns (A)[NO-NO] and (B)[NO-YES], which are the relevant ones to test the moral law, one arrives at the final two columns, which represent the amount of partial confirmation (A')[NO-NO] and the amount of disconfirmation (B')[NO-YES] to the moral law in each of the pairs of scenarios of the replication.

In order to eliminate the harm sceptics' pooling problem, I shall make an educated guess. If one takes the replication results of the blatantly counter-intuitive pattern (D)[YES-NO] as an estimate of the harm sceptics' results of pattern (D)[YES-NO], that is, if one supposes that in each of their pairs of scenarios the percentage of column (D) is close to 0% (0%, to simplify), it is a question of trivial combinatorial logic and arithmetic to arrive at a good estimate of their missing results. If pattern (D)[YES-NO] is 0%, then, in each pair, the percentage of column (C) is the same as the percentage of column (1). This is because (1) represents all participants who said YES to the first scenario, there are only two ways of answering in this way, (C)[YES-YES] and (D)[YES-NO], and (D)[YES-NO] is being assumed to be 0%. For example, in their slavery pair, column (C) becomes 7%. Since the percentages in column (2) represent all participants who said YES to the second scenario in each pair, and there are only two ways of doing that, (C)[YES-YES] and (B)[NO-YES], one of which is now known, by subtracting in each pair the percentage of column (C) from the percentage of column (2), one gets the percentage of column (B). In their slavery pair, the percentage of column (B) becomes 4% (11% minus 7%). In each pair, the percentage of pattern (A)[NO-NO] is the remaining one, which also corresponds to 100% minus the percentage of column (2), for similar deductive reasons. In the slavery pair, column (A) becomes 89%. In a nutshell, in Table 2, if (D) is 0%, then (C) equals (1), (B) equals (2) minus (1), and (A) equals 100% minus (2). Now, to arrive at the percentages of (A') and (B') in each of their pairs, one has first to calculate the amount of participants in each of the columns (A), (B) and (C) of each pair from their percentages and the total number of participants (N), and then to recalculate the percentages of (A) and (B) by excluding the number of participants in (C) from the total. The final result of their slavery pair becomes

95% of partial confirmation (A')[NO-NO] and 5% of disconfirmation (B')[NO-YES] of the moral law.

As one can see in columns (A') and (B'), except for the whipping generality pair, the results of the replication are quite similar to the harm sceptics' reanalyzed results. However, these results do not strike one as the substantial evidence against the moral law that the harm sceptics, based on the pooled data described in columns (1) and (2), claim. To give the most extreme example, take what they say about the slavery scenarios: 'In the Slavery scenarios ..., we again found a *dramatic difference* ... 11% of subjects reported that slavery was OK in Greco-Roman societies, but only 7% reported that it was OK in the American South ($p = 0.021$)' (Kelly *et al.* 2007, p. 126; my emphasis). Now, having the data properly described to eliminate the pooling and permissibility problems, the results show 95% (100% in the replication) of partial confirmation and 5% (0% in the replication) of disconfirmation of the moral law, a completely different picture.

It is true that the results in column (B') still show a sizeable amount of participants whose pattern of response disconfirms the moral law, especially in the whipping generality pair (harm sceptics' results) and in the training authority pair (both results). The next section questions whether the responses of these participants really reflect disconfirmation of the moral law.

4. The Confounding Variable Problem

The problem of this section has to do with the fact that the OK questions of the moral/conventional task do not necessarily lead participants to make judgements of wrongdoing because the OK question is polysemous. First, I characterize the problem; then, I draw a possible consequence that would weaken the evidence against the moral law even more.

In the standard moral/conventional task, the OK question constitutes the permissibility probe *and* is part of the authority contingency and generality probes. The grammatical form of the OK question is represented in (0), dealing here only with Hs.

(0) Is it OK for X to H? YES NO

The intended meaning of the question asks participants to make an *evaluative judgement*. If the question is understood accordingly, when a participant answers YES, she is saying that, in doing H, X did not do something wrong, while, when she answers NO, she is saying that, in doing H, X did something wrong. In both cases, the participant is evaluating H with her judgement. I will represent this evaluative meaning by (1).

(1) *Is it OK that X does H?*

There is another reading of (0) though, simply asking participants to make a *non-evaluative description*. This descriptive meaning is represented by (2).

(2) *According to X, is it OK that X does H?*

If (0) is understood as (2), when a participant answers YES, he is saying that, according to X, in doing H, X did not do something wrong, while when a participant answers NO, he is saying that, according to X, in doing H, X did something wrong. In both cases, the participant herself is not making an evaluative judgement; she is just describing X's evaluative judgement of H.

A relevant variant of reading (2) is represented by (3), where 'Z' refers to persons other than X (with the caveat that it should not refer to the participant of the task) or to more abstract entities such as the legal system.

(3) *According to Z, is it OK that X does H?*

Again, if (0) is understood as (3), in answering YES or NO, the participant herself is not making an evaluative judgement—she is just describing Z's evaluative judgement of H.

As I mentioned in the last section, in the replication of the harm sceptics' research, participants were asked to *justify their YES/NO answer to the OK questions*—a justification probe was introduced in their design. The three different readings of the OK question just discussed are envisaged in the following answer to the justification probe of the first scenario of the training authority pair: 'It is not OK for him to do it in the context of his orders and military law [reading (3)]. Whether it is OK for him to do it according to his own morals I can't say [reading (2)] but I still think its OK [reading (1)]'.¹⁰

Interestingly, the NO answer related to this explanation shows that the participant *did not* understand the OK question in terms of the evaluative reading (1). Here is the participant's complete response:

TRAINING AUTHORITY PAIR (nonstandard order¹¹)

SCENARIO 1 (physical abuse forbidden by authority and law)

OK question: Is it OK for Sergeant Anderson to physically abuse trainees during simulated interrogations?

Answer: NO

¹⁰ It is fundamental to distinguish the overall answer given to the justification probe from its part that is intended to justify the YES/NO answer to the OK question: this participant is not using all that is said here to justify the YES/NO answer. I utilize the word 'justification' only in the sense of *justification for the YES/NO answer to the OK question*, and the word 'explanation' in the sense of *overall answer to the justification probe*.

¹¹ 'Nonstandard order' means that scenario 1 was presented to the participant after scenario 2, while 'standard order' means that scenario 1 was presented first.

Explanation: It is not OK for him to do it in the context of his orders and military law. Whether it is OK for him to do it according to his own morals I can't say but I still think its OK.

SCENARIO 2 (physical abuse permitted by authority and law)

OK question: Is it OK for Sergeant Anderson to physically abuse trainees during simulated interrogations?

Answer: YES

Explanation: The commandos are being trained for combat, and capture and interrogation is a part of war. They must know what to expect and be trained to deal with it. Rugby players often get bruised during training sessions, this is not permanent damage and probably makes them stronger.

In this [NO-YES] pattern of response, the NO answer is justified by the passage 'It is not OK for him to do it in the context of his orders and military law', which evinces a descriptive reading of the OK question, not by the passage 'I still think its OK', which would evince an evaluative reading of the OK question; the YES answer seems to *endorse* the point of view of the authority/law, since its justification is similar to the rationale given by the military, and the subsequent explanation says 'I *still* think its OK' (see note 11). If my interpretation is correct, while in scenario 1 there was a descriptive reading of the OK question, in scenario 2 there was an evaluative reading, since endorsements of (or, more generally, agreements with) evaluations are evaluations.

Here is another example of a [NO-YES] pattern of response:

WHIPPING GENERALITY PAIR (standard order)

SCENARIO 1 (to whip forbidden now)

OK question: Is it OK for Mr. Adams to whip the sailor?

Answer: NO

Explanation: From modern standards I believe receiving lashes for insubordination is too extreme.

SCENARIO 2 (to whip permitted 300 years ago)

OK question: Is it OK for Mr. Williams to whip the sailor?

Answer: YES

Explanation: Such was the standard of the time and what was expected, from the sailors and the officers. However it is still wrong.

In this [NO-YES] pattern of response, the justification to the NO answer seems to imply agreement with the evaluation coming from modern standards, since the subsequent explanation says 'however, it is *still* wrong'; the YES answer is justified by the passage 'Such was the standard of the time and what was expected ...', which evinces a descriptive reading of the OK question, not by the passage 'however it is still wrong', which would evince an evaluative reading of the OK

question. If my interpretation is correct, again there is a shift in the reading of the OK question, now from evaluative to descriptive.¹²

Let me now give two examples of a [NO-NO] pattern of response:

TRAINING AUTHORITY PAIR (nonstandard order)

SCENARIO 1 (physical abuse forbidden by authority and law)

OK question: Is it OK for Sergeant Anderson to physically abuse trainees during simulated interrogations?

Answer: NO

Explanation: Well, I marked it as not OK in the first place, and it still isn't okay here. In fact now it is doubly wrong—it is in violation of the official law of the government and his supervisors. It should still be wrong with his personal values as well.

SCENARIO 2 (physical abuse permitted by authority and law)

OK question: Is it OK for Sergeant Anderson to physically abuse trainees during simulated interrogations?

Answer: NO

Explanation: It might be okay legally—after all, he is following the orders of his commanders. However, threats and abuse are generally not OK. Even if they have data proving the training was helpful (something more than 'belief' would be necessary), they must be sure they aren't doing lasting damage to the recruits.

WHIPPING GENERALITY PAIR (standard order)

SCENARIO 1 (to whip forbidden now)

OK question: Is it OK for Mr. Adams to whip the sailor?

Answer: NO

Explanation: why should it be?

SCENARIO 2 (to whip permitted 300 years ago)

¹² We don't know a priori the way participants understand the question. We have to count on their explanations to make reasonable hypotheses thereof. The problem is that *statements* like 'according to X (Z), it is Not-OK/OK' do not necessarily indicate that a participant intends to convey simply a description—one may be building an argument of authority to support one's own implicit evaluation (e.g. in saying 'according to God, this is Not-OK', one may be implicitly affirming the evaluative conclusion 'therefore, this is Not-OK'). On the other hand, *statements* like 'it is Not-OK/OK' do not necessarily indicate that a participant intends to convey an evaluation—one may be using some kind of free indirect style, leaving implicit an 'according to X (Z)'. In my interpretation, I'm taking 'however it is still wrong' as conveying an evaluation and, for this reason, taking 'From modern standards I believe receiving lashes for insubordination is too extreme' as implying some evaluation. That's why I think there is an evaluative reading of the OK question in scenario 1. Alternatively, one may take 'however it is still wrong' in terms of a free indirect style—'however it is still wrong (from modern standards)'. This may lead one to claim that in fact the reading of the Ok question in scenario 1 is simply descriptive, with no shift involved. I think 'however' favours my interpretation, though it is probably the case that, in scenario 1, the participant was somewhat confused by the two possible readings of the OK question.

OK question: Is it OK for Mr. Williams to whip the sailor?

Answer: NO

Explanation: It depends what OK means ... If it means if it is acting according to the rules, then it is OK ... If it means it is ok according to [our] rules ... Probably the answer is no, we have different ways of whipping nowadays. We don't actually whip, but we use other means. If it means that it was the practice that worked at that time ... Probably the answer is ambiguous. I find the question too vague ... Anyway I am against whipping so I answer no.

In these [NO-NO] patterns of response, the justifications seem to evince that all NO answers were driven by an evaluative reading of the OK question.

The explanations of these four participants confirm the polysemy of the OK question, which introduces a confounding variable that may compromise both the validity and internal reliability of the dependent measure of the moral/conventional task. This is the confounding variable problem. Before moving on, I would like to point out two things. This type of confounding variable would not be eliminated in the design versions where the word 'wrong' substitutes for the word 'OK' in the questions. Throughout this article, I've been supposing that the intended meaning of the OK question is the evaluative one I characterized, but authors who make usage of the moral/conventional task are not always perspicuous in this respect; whichever the intended meaning, inasmuch as the confounding variable remains, the problem persists.

From this problem, I draw now a possible consequence for the results of the harm sceptics as far as the test of the moral law is concerned. Because their design does not include a justification probe, one cannot identify the problematic data points of their results, that is, the ones where the OK question is not properly understood. I can use the explanations of the replication to suggest a general correction that would weaken the evidence against the moral law even more.

Notice that only the problematic participants described above, the first two, exhibit the pattern (B)[NO-YES] of response, which is the one that disconfirms the moral law (see previous section). Notice also that their two pairs correspond to those that gave most disconfirmation to the moral law [see columns (A') and (B') of whipping generality and training authority in Table 2]. The problematic character of these two participant's responses is already sufficient reason to eliminate both of them from the analysis, diminishing thereby the disconfirmation of the moral law. But now, if one were to use these two participants' explanations to discover their YES/NO answers to the evaluative reading of the OK question, the first participant would show the pattern (C)[YES-YES] and the second the pattern (A)[NO-NO]. The former, being irrelevant to test the moral law, would be eliminated from the analysis, thus diminishing the disconfirmation of the moral law, whereas the latter, being relevant, would be included, thus increasing the confirmation of the moral law.

More generally, in most of the harm sceptics' paired scenarios, because there is a Not-OK/OK asymmetry between the first and the second scenarios according to

the institutionalized context, a descriptive reading of the OK questions tends to lead to a pattern (B)[NO-YES] that does not reflect a real disconfirmation of the moral law. In the slavery pair, where such an asymmetry does not exist, since its scenarios describe harmful actions that are OK according to the institutionalized context, a descriptive reading of the OK question would lead to a pattern (C)[YES-YES] instead. This rationale is somewhat confirmed by the fact that the only participant in the replication slavery pair with the pattern (C)[YES-YES] seemed to be guided by a descriptive reading of the two OK questions:¹³

SLAVERY GENERALITY PAIR (standard order)

SCENARIO 1 (to keep slaves permitted—American South)

OK question: Was it OK for Americans to keep slaves?

Answer: YES

Explanation: Because as stated above they were an important part of the economy and their position in society was considered 'normal' at the time. Although one could argue that even in the 19th century there were people calling for the abolition of slavery in the US, so maybe even then they understood that something abnormal stood beyond slavery.

SCENARIO 2 (to keep slaves permitted—Greece and Rome)

OK question: Was it OK for the ancient Greeks and Romans to keep slaves?

Answer: YES

Explanation: It was considered normal for that society to have slaves since they had different notions of human rights and citizenship.

I believe my general suggestion goes in the right direction indeed. The next section deals with another fashion in which participants' YES/NO patterns of response in the context of the harm sceptics' scenarios may be misleading.

5. The Law Scope Problem

By the end of their article, the harm sceptics raise some questions to be addressed in future research, including: '... why did previous research on schoolyard harm

¹³ Alternatively, one may envisage this participant having had an evaluative reading of the OK question, by interpreting participant's justifications as implying *evaluative relativist* YES answers (it's OK because OK-ness is relative), instead of *descriptive relativist* ones (according to X, it is OK). A descriptive relativist claim is just a description of other people's *different* evaluations without implying anything about whether their evaluations are valid. An evaluative relativist claim accepts the validity of other people's *different* evaluations, but just because validity is a relative matter. This evaluative relativist interpretation would not support the moral law in terms of the way the harm sceptics originally envisaged these scenarios, although it would support it if one were to take both scenarios as involving distant past and were to assume that the participant thinks that slavery is Not-OK now (see note 7). I think the passage 'maybe even then they understood that something abnormal stood beyond slavery', which seems to imply the *tout court* non-validity of slavery, supports my interpretation.

transgressions appear to support (C-2a) [the moral law]? Is there something special about these simple harm transgressions that is not shared by the more “grown-up” transgressions that we also used in our study?’ (Kelly *et al.*, 2007, p. 129) I have argued that, in terms of support for the moral law, their ‘grown-up’ results are not as dissimilar as they suppose. I agree that there is a crucial difference between the Turiel tradition’s harm scenarios and theirs, but one that in my opinion leads to the problem of this section. By explicating the crucial difference, I show the final problem with its consequence to the test of the moral law.

The moral law says *If transgression H, then moral signature*. In the context of the Turiel tradition, H involves a victim being harmed *and* being subjected to injustice/rights violation. Accordingly, to categorize an action as an H is tantamount to considering it a transgression—the injustice/rights violation related to the doing implies wrongdoing. Again, there is no logical or empirical necessity that what researchers classify as H will be categorized by participants as a transgression, even if a harmful action is explicitly depicted as an instance of injustice/rights violation, since participants need not agree with the point of view of the depiction. However, the ‘schoolyard’ harm scenarios of the Turiel tradition involve prototypical cases where the link between harm and injustice/rights violation (and hence transgression) is the most obvious and consensual interpretation. That’s why, in comparison with Ns, Hs are rarely considered OK in the ‘schoolyard results’ of the permissibility probe of the standard moral/conventional task (see section 2). In other words, the categorization of these prototypical actions matches the antecedent of the moral law.

To understand the singular character of the harm sceptics’ scenarios, it is important to notice that the general notion of harm as pain or suffering is necessarily linked neither to injustice/rights violation, nor to transgression. Harm in this general sense, instead of the more specific sense referenced by ‘H’, may be perceived as permissible (e.g. the pain or suffering caused in self-defence or in the context of certain types of trolley dilemmas) and even as obligatory (e.g. the pain or suffering caused by deserved punishment, by some types of medical or war treatment, or by some types of training procedures). Now, in most of the harm sceptics’ paired scenarios, harm is portrayed in the complex contexts of punishment, war treatment, and professional training, just the types of contexts that leave the deontic status of harm more open to different interpretations. This is one of the reasons why an important number of participants answer OK to the permissibility probe in the first scenarios of the harm sceptics’ pairs. For example, take the explanations to the first scenarios given by these participants who showed a pattern (C)[YES-YES]:

WHIPPING AUTHORITY PAIR (nonstandard order)

SCENARIO 1 (whipping forbidden by authority)

OK question: Is it OK for Mr. Adams to whip the sailor?

Answer: YES

Explanation: He won’t get drunk again. He could have risked the lives of his sea mates.

TRAINING AUTHORITY PAIR (standard order)

SCENARIO 1 (physical abuse forbidden by authority)

OK question: Is it OK for Sergeant Anderson to physically abuse trainees during simulated interrogations?

Answer: YES

Explanation: They need to be ready for what they will be faced with in war.

The first participant justifies the deserved punishment by appealing to its deterrent function (and perhaps by evoking the culpability of the sailor given the possible consequences of his negligent behaviour), while the second justifies the propriety of the training by evoking some kind of precautionary reasoning. In both, the pain or suffering involved is accepted by appealing to a rationale that breaks the connection between harm *and* injustice/rights violation existent in the antecedent of the moral law.

But these kinds of contexts have implications to the moral *signature* as well. Take the explanations to the second scenarios of the pairs given by these participants who showed a pattern (B)[NO-YES], which apparently would disconfirm the moral law:

WHIPPING AUTHORITY PAIR (standard order)

SCENARIO 2 (whipping permitted by authority)

OK question: Is it OK for Mr. Adams to whip the sailor?

Answer: YES

Explanation: As long as Mr. Adams [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.

PRISONER AUTHORITY PAIR (nonstandard order)

SCENARIO 2 (physical abuse permitted by authority)

OK question: Is it OK for Sergeant Johnson to keep the suspect awake for three days and three nights?

Answer: YES

Explanation: If they have reason enough to suspect him in the first place then it makes sense.

The first participant seems to justify the punishment by the fact that the wrongdoer was aware of the consequences of his actions, while the second seems to justify the treatment by the possibility of obtaining reliable information. The participants do not seem to be appealing to *the authority of the superiors in itself* to justify the OK of the punishment and treatment. Therefore, because their YES answers do not seem to *negate* the moral signature's authority conjunct (*Not-OK even when an authority says OK*), their pattern (B)[NO-YES] may not reflect a disconfirmation of the moral law.

Take now the explanations to the second scenarios of the pairs given by these participants who showed a pattern (A)[NO-NO], which apparently would give partial confirmation to the moral law:

TRAINING AUTHORITY PAIR (standard order)

SCENARIO 2 (physical abuse permitted by authority)

OK question: Is it OK for Sergeant Anderson to physically abuse trainees during simulated interrogations?

Answer: NO

Explanation: The statement did not say that it was helpful for sure. It stated that they 'thought' it was helpful. I think there have to be ways to prepare soldiers for what they will face without out right physical abuse as described in this study.

PRISONER AUTHORITY PAIR (standard order)

SCENARIO 2 (physical abuse permitted by authority)

OK question: Is it OK for Sergeant Johnson to keep the suspect awake for three days and three nights?

Answer: NO

Explanation: The suspected will be confused and therefore may answer questions inaccurately.

The first participant seems to deny the propriety of the training procedure primarily due to its uncertain results, while the second denies the propriety of the treatment due to its certain inefficacy. Neither of them seems to ground their NO answers on issues of harm and injustice and/or rights violation that would eliminate the normative effectiveness of an authority decision. Therefore, because their NO answers do not really seem to affirm the moral signature's authority conjunct (*Not-OK even when an authority says OK*), their pattern (B)[NO-NO] may not constitute partial confirmation of the moral law.

In sum, the crucial difference between the schoolyard harm scenarios of the Turiel tradition and those of the harm sceptics is that the contexts of the latter activates kinds of reasoning that go beyond the scope of the moral law. The law scope problem is that, even when the YES/NO patterns of response relevant to test the moral law seem to be driven by the evaluative reading of the OK question, they may not test it, given that the reasoning behind them may be unrelated to the scope of the moral law.

6. Conclusion

One may have perceived some irony in my discussion throughout the article. It is as if in the process of defending the moral law, I have been undermining the Turiel tradition. One may even think that my perspective embeds the following paradox: my meta-scepticism towards the harm sceptics' scepticism towards the Turiel tradition actually fortifies their scepticism. I conclude by clarifying the import of my meta-scepticism.

The harm sceptics' critical approach to the Turiel tradition supposes that one may isolate different facets of this tradition—the moral law, the conventional law... In this article, not only have I accepted this supposition but also have pursued it. Accordingly, this is my way out of the paradox: I believe that my discussion undermines the harm sceptics' scepticism towards the moral law favouring thereby one aspect of the Turiel tradition, although I also believe that my discussion may have critical implications to other aspects of this tradition, in particular, to the moral/conventional distinction.

My discussion of the permissibility problem favours the moral law, since the elimination of participants who answered OK to the permissibility probe increases its confirmation [cf. (A)/(B) and (A')/(B') in Table 2].

The permissibility problem may have a critical consequence for the moral/convention distinction, though. Imagine a participant answering YES both to the permissibility probe 'Is it OK for X to N (e.g. *a child eating lunch with fingers*)?' and to the authority contingency probe 'Now, what if an authority says that it is OK to N. Would it be OK for X to N?'¹⁴ If this [YES-YES] sequence of data points is not eliminated from the final analysis, the second YES will be pooled with the legitimate YES answers to the authority contingency probe (i.e. those coming from a [NO-YES] sequence) to constitute the total (aggregate level) of YES answers to the authority contingency probe related to Ns. Then, this total will be compared with the total of YES answers to the authority contingency probe related to Hs. Since the YES answer to the permissibility probe (and therefore also the sequence [YES-YES]) is more frequent on the conventional side of the standard moral/conventional task, in including irrelevant data, one is increasing the pooled difference between convention and morality in the authority contingency probe. The danger of this inflation of difference is the possibility that some of the established statistically significant results of the current literature are in fact an overestimation of the moral/convention distinction.¹⁵

My discussion of the confounding variable problem favours the moral law, since, as I suggested, this is really a problem only for the pattern of results that disconfirm it—i.e. [NO-YES]. One of the anonymous referees raised the possibility of the polysemy of the OK question being equally damaging to the pattern of results that confirms the moral law—i.e. [NO-NO]. It may be that participants are showing this pattern based on a descriptive reading of the OK question—e.g. answering Not-OK according to current norms (to the permissibility probe)

¹⁴ The pattern [YES-NO] is counterintuitive here as well—see discussion of column (D) of Table 2.

¹⁵ The other data analysis problem, the pooling problem, has no critical consequence to the moral/conventional distinction. Leaving aside the permissibility problem, the pooling of the data described in this paragraph, contrary to the pooling perpetrated by the harm sceptics, is perfectly legitimate. Perhaps the harm sceptics ran into their puzzling pooling problem because they made a superficial analogy between data analysis in the context of the standard moral/conventional task, which deals with the moral/convention distinction, and data analysis in the context of testing the moral law in itself.

and Not-OK according to current norms (to the generality probe) by following simply a descriptive reading. Although I cannot address this possibility in detail here, let me give two reasons coming from the data of the replication that, albeit not decisive, make this prospect unattractive. The first is just a more general reiteration of the type of evidence supporting my suggestion in section 4: the cases where participants *explicitly acknowledge the polysemy in their explanations*, a sample of which was presented in section 4, corroborate my suggestion. The second is related to the interpretation of descriptive statements like ‘according to X, it is Not-OK’ (see also note 12). Participants did make such statements when explaining their [NO-NO] answers. For example, in the slavery scenarios, a participant would make a justification like ‘according to current notions of human rights, it is wrong’. However, if one takes the overall explanation of the participant into account, the best hypothesis seems almost always to be that an evaluation by means of an endorsement or an argument of authority is being implied by such type of statement, which indicates an evaluative reading of the OK question. For example, the overall explanation of the participant would *accumulate* other justifications such as ‘it is cruel’, ‘it is inhumane’ that seem plainly evaluative.

The confounding variable problem may have a critical consequence for the moral/conventional distinction, though. Reading shifts or descriptive readings may be even more accentuated in conventional cases. If this is so, the problem here is that part of the asymmetry of results in the standard moral/conventional task may be due to the fact that whereas participants tend to have an evaluative reading of the OK questions in cases of H, they tend to have a descriptive reading in cases of N. I’m currently running research to test this hypothesis.¹⁶

My discussion of the law scope problem favours the moral law by undermining the simplistic rationale behind the harm sceptics’ attempt to disconfirm it. One should notice that some harmful actions may invoke types of reasoning unrelated to the moral law and that, while these actions do so, they are irrelevant to test it. One specific objection raised by the same anonymous referee is that with my discussion of the scope problem, the moral law starts to look like an utterly ad hoc hypothesis: isn’t it the case that only harmful actions that confirm the moral law are supposed to test it? A way of addressing this specific issue is by posing another question: what is the relevant concept of harmful action that could be part of an interesting hypothesis about how people understand the relation between harm and moral wrongdoing? In other words, what is the appropriate characterization of the concept of harmful action as part of the antecedent of the moral law? It cannot simply be the concept of *an action causing pain or suffering*, since this would include too many cases of permitted and even obligatory actions, and the topic at stake is the concept of moral *wrongdoing*. Consequently, the concept has to be at least thus

¹⁶ Notice that, in introducing two justification probes in the design of the harm sceptics, the replication obtained a type of data not provided by the standard moral/conventional task, which does not include justification probes related to the authority and generality probes.

specific—a *transgression causing pain or suffering*. Now, can this more specific concept include, besides cases that involve injustice/rights violation, cases that do not involve injustice/rights violation? If no, this concept coincides with the concept of H. If yes, wouldn't these additional cases be rather related to domains of deontic reasoning, like prudential or precautionary reasoning, that are to be considered outside of the scope of *moral wrongdoing*? I don't have a clear answer to these last two questions, but it looks like that, after all, the interesting hypothesis on the understanding of the relation between harm and moral wrongdoing is indeed the moral law—if *transgression H*, then *moral signature*. But then, it is not that only harmful actions that confirm the moral law are supposed to test it, it is just that, to test it, one has necessarily to deal with cases interpreted by participants as H.

My position on the scope of the moral law may have a critical consequence for (strict interpretations of) the moral/conventional distinction, though. It accepts that the Turiel tradition is on the right track as far as the folk understanding of the relation between harm and moral wrongdoing is concerned, but without necessarily implying that the moral law exhausts the moral-wrongdoing domain. In the way I framed the moral law, Hs are 'sufficient but not necessary conditions'. This leaves open the possibility that certain *transgressions not causing pain or suffering* be conceived as part of morality, and that's how I would like to close my discussion.

*Institute of Cognition and Culture
Department of Anthropology
Queen's University, Belfast*

References

- Blair, R. 1995: A cognitive developmental approach to morality: investigating the psychopath. *Cognition*, 57, 1–29.
- Blair, R. 1996: Brief report: morality in the autistic child. *Journal of Autism and Developmental Disorders*, 26, 571–579.
- Blair, R. 1997: Moral reasoning and the child with psychopathic tendencies. *Personality and Individual Differences*, 26, 731–739.
- Fiddick, L. 2004: Domains of deontic reasoning: resolving the discrepancy between the cognitive and moral reasoning literatures. *The Quarterly Journal of Experimental Psychology*, 57, 447–474.
- Gabennesch, H. 1990: The perception of social conventionality by children and adults. *Child Development*, 61, 2047–2059.
- Grice, P. 1991 [1968]: Logic and conversation. In S. Davis (ed.), *Pragmatics—a Reader*. Oxford: Oxford University Press.
- Haidt, J., Koler, S. and Dias, M. 1993: Affect, culture and morality, or is it wrong to eat your dog? *Journal of Personality and Social Psychology*, 65, 613–628.
- Kelly, D. and Stich, S. (in press): Two theories about the cognitive architecture underlying morality. In P. Carruthers, S. Laurence and S. Stich (eds), *The Innate Mind*, Vol. III, *Foundations and the Future*. Oxford: Oxford University Press.

- Kelly, D., Stich, S., Haley, K., Eng, S. and Fessler, D. 2007: Harm, affect, and the moral/conventional distinction. *Mind & Language*, 22, 117–131.
- Nichols, S. 2002: Norms with feeling: toward a psychological account of moral judgement. *Cognition*, 84, 223–236.
- Nichols, S. 2004: *Sentimental Rules: On the Natural Foundations of Moral Judgement*. Oxford: Oxford University Press.
- Nucci, L. 1981: Conceptions of personal issues: a domain distinct from moral or societal concepts. *Child Development*, 52, 114–121.
- Nucci, L. 2001: *Education in the Moral Domain*. Cambridge: Cambridge University Press.
- Nucci, L. and Herman S. 1982: Behavioural disordered children's conceptions of moral, conventional, and personal Issues. *Journal of Abnormal Child Psychology*, 10, 411–425.
- Nucci, L. and Turiel, E. 1978: Social interactions and the development of social concepts in preschool children. *Child Development*, 49, 400–407.
- Smetana, J. 1981: Preschool children's conceptions of moral and social rules. *Child Development*, 52, 1333–1336.
- Smetana, J. 1985: Preschool children's conceptions of transgressions: effects of varying moral and conventional domain-related attributes. *Developmental Psychology*, 21, 18–29.
- Smetana, J. 1986: Preschool children's conceptions of sex-role transgressions. *Child Development*, 57, 862–871.
- Smetana, J. 1993: Understanding of social rules. In M. Bennett (ed.), *The Development of Social Cognition: The Child as Psychologist*. New York: Guilford Press.
- Smetana, J. and Braeges, J. 1990: The development of toddlers' moral and conventional judgements. *Merrill-Palmer Quarterly*, 36, 329–346.
- Smetana, J., Schlagman, N. and Adams, P. 1993: Preschool children's judgements about hypothetical and actual transgressions. *Child Development*, 64, 202–214.
- Smetana, J., Kelly, M. and Twentyman, C. 1984: Abused, neglected, and nonmaltreated children's conceptions of moral and social-conventional transgressions. *Child Development*, 55, 277–287.
- Sousa, P., Holbrook, C. and Piazza, J. forthcoming: The morality of harm. *Cognition*.
- Tisak, M. 1995: Domains of social reasoning and beyond. In R. Vasta (ed.), *Annals of Child Development*, Vol. 11. London: Jessica Kingsley.
- Tisak, M. and 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.
- Weston, D. and Turiel, E. 1980: Act-rule relations: children's concepts of social rules. *Developmental Psychology*, 16, 417–424.
- Yau, J. and Smetana J. 2003: Conceptions of moral, social -conventional, and personal events among Chinese preschoolers in Hong Kong. *Child Development*, 74, 647–658.