Personality and Social Psychology Bulletin
2015, Vol. 41(5) 696–713

DOI: 10.1177/0146167215575731
pspb.sagepub.com
SAGE

Gender Differences in Responses to Moral Dilemmas: A Process Dissociation Analysis

Rebecca Friesdorf[1], Paul Conway[2], and Bertram Gawronski[3]

Abstract

The principle of *deontology* states that the morality of an action depends on its consistency with moral norms; the principle of *utilitarianism* implies that the morality of an action depends on its consequences. Previous research suggests that deontological judgments are shaped by affective processes, whereas utilitarian judgments are guided by cognitive processes. The current research used process dissociation (PD) to independently assess deontological and utilitarian inclinations in women and men. A meta-analytic re-analysis of 40 studies with 6,100 participants indicated that men showed a stronger preference for utilitarian over deontological judgments than women when the two principles implied conflicting decisions (d = 0.52). PD further revealed that women exhibited stronger deontological inclinations than men (d = 0.57), while men exhibited only slightly stronger utilitarian inclinations than women (d = 0.10). The findings suggest that gender differences in moral dilemma judgments are due to differences in affective responses to harm rather than cognitive evaluations of outcomes.

Keywords

gender, meta-analysis, moral dilemmas, moral judgment, process dissociation

Received May 13, 2014; revision accepted February 9, 2015

> Men are governed by lines of intellect—women: by curves of emotion.
>
> —James Joyce (1882-1941)

> I hate to hear you talk about all women as if they were fine ladies instead of rational creatures.
>
> —Jane Austen (1775-1817)

The similarities and differences between men and women have fascinated intellectuals and lay people alike. Many, including author James Joyce, have argued for the old stereotype that men engage in more cognitive activity, whereas women experience greater affectivity. Yet others, like Jane Austin, have argued that gender differences are exaggerated. In the field of psychology, gender differences have been investigated in many domains (e.g., Parsons & Bales, 1955), and the domain of morality is no exception. Gilligan (1982) famously argued that male morality is relatively more cognitive and depersonalized than female morality, which is relatively more affect-laden and personalized.

Although the evidence regarding gender differences in moral reasoning is mixed (Jaffee & Hyde, 2000), recent work suggests systematic gender differences in responses to moral dilemmas where causing some degree of harm leads to greater well-being overall (e.g., Fumagalli et al., 2010). However, the available evidence remains ambiguous regarding whether these differences are driven by gender differences in cognitive evaluations of action outcomes, affective responses to harmful actions, or a combination thereof. The current work adopts a process dissociation (PD) approach (Conway & Gawronski, 2013b) to clarify the nature of gender differences in responses to moral dilemmas.

Gender Differences in Moral Reasoning

For years, the field of moral psychology was dominated by the rationalist approach of Lawrence Kohlberg, who described a trajectory of moral development toward the rational application of abstract moral principles to resolve moral conflicts (e.g., Kohlberg, 1969). Carol Gilligan (1982), a student of Kohlberg, criticized his approach as overly male-centric. She argued that men prefer a cognitive, abstract, depersonalized method of moral decision making,

[1]Wilfrid Laurier University, Waterloo, Ontario, Canada
[2]University of Cologne, Cologne, Germany
[3]University of Texas at Austin, USA

Corresponding Author:
Rebecca Friesdorf, Department of Psychology, Wilfrid Laurier University, 75 University Avenue W, Waterloo, Ontario, Canada N2L 3C5.
Email: frie3750@mylaurier.ca

which she dubbed an *ethic of justice*. Women, in contrast, prefer an *ethic of care*, involving moral decisions based on relations to and emotional bonds with particularized others. Hence, in Gilligan's view, men typically make moral decisions from a detached, individualistic, third-party perspective. Conversely, women typically make moral decisions from an emotionally engaged, interpersonal, first-person perspective (Benhabib, 1985).

Gilligan's gender theory of morality proved popular and stimulated a great deal of research over the ensuing years. Some studies supported Gilligan's view (e.g., Gilligan & Attanucci, 1988; Johnston, 1988), whereas others did not (e.g., Krebs, Vermeulen, Denton & Carpendale, 1994; Skoe & Diessner, 1994). Ultimately, Jaffee and Hyde (2000) conducted a meta-analysis examining 113 studies and drew the conclusion that gender differences in moral orientations were small and often non-significant. These modest results quelled further research on this question.

However, a lack of gender differences in care and justice orientations should not be equated with a general lack of gender differences in moral psychology. After all, gender differences in affective and cognitive processing may influence moral judgments in a manner that does not map onto the two kinds of ethics proposed by Gilligan. This possibility is consistent with newer conceptions that emphasize the distinct contributions of affective and cognitive processes to moral decision making (Greene & Haidt, 2002).

Dual-Process Model of Moral Judgment

One of the most prominent examples of such conceptions is Greene's (2007) dual-process model of moral judgment, which has its roots in philosophical conundrums where causing some harm leads to a better overall outcome. Consider, for example, a dilemma where you are hiding with other townsfolk from murderous soldiers. Suddenly a baby starts to cry—unless you smother it, the soldiers will find and kill everyone. Should you smother the baby to prevent the soldiers from killing the townsfolk? Such dilemmas are said to pit deontological against utilitarian ethical positions (Foot, 1967). According to the principle of *deontology*, killing the baby is morally wrong because the morality of an action hinges on its consistency with context-independent moral norms, in this case the norm not to kill others (Kant, 1785/1959). According to the principle of *utilitarianism*, killing the baby is acceptable because the morality of an action hinges on its overall consequences, in this case the many lives that will be saved (Mill, 1861/1998).

Although originally conceived of as philosophical problems, Greene, Sommerville, Nystrom, Darley, and Cohen (2001) argued that moral dilemma judgments are influenced by two psychological processes: (a) a relatively affect-laden reaction to the idea of harming a specific individual and (b) a relatively cognitive evaluation of outcomes. Whereas the former motivates harm rejection regardless of the outcomes (consistent with the principle of deontology), the latter motivates harm acceptance when harm maximizes overall outcomes (consistent with the principle of utilitarianism). A large body of evidence supports the view that rejecting harmful action in moral dilemmas is associated with affective processes, and accepting harmful action that maximizes outcomes is associated with cognitive processes. For example, manipulations that vividly highlighted the harm caused by action increased deontological judgments (Amit & Greene, 2012; Bartels, 2008), whereas reducing negative affect reduced deontological judgments (Strohminger, Lewis, & Meyer, 2011). Conversely, manipulations enhancing rational decision making increased utilitarian judgments (Bartels, 2008; Nichols & Mallon, 2006), whereas cognitive load slowed (Greene, Morelli, Lowenberg, Nystrom, & Cohen, 2008) and time pressure reduced (Suter & Hertwig, 2011) utilitarian judgments. Moreover, brain regions associated with affect were more active when people made deontological judgments, whereas brain regions associated with working memory were more active when participants made utilitarian judgments (Greene, Nystrom, Engell, Darley, & Cohen, 2004). Thus, a wide range of evidence supports the dual-process model of moral judgment where affective reactions to harm motivate harm rejection regardless of the outcomes (consistent with deontological ethics), and cognitive evaluations of outcomes motivate harm acceptance when harm leads to better overall outcomes (consistent with utilitarian ethics).

Gender Differences From a Dual-Process Perspective

Despite the weak evidence for gender differences in care and justice orientations, Greene's (2007) dual-process model offers a novel perspective on gender differences in moral psychology on the basis of fundamental differences in non-moral information processing. Although there is little evidence for gender differences in cognitive processing, gender differences in affective processing are common and robust. For example, men and women score equally high on need for cognition (Cacioppo, Petty, Feinstein, & Jarvis, 1996), and gender differences in cognitive ability tend to be rather small (Hyde, 1981). Yet, women tend to experience stronger emotional responses than men (e.g., Brody & Hall, 2000; Cross & Madson, 1997; Fischer & Manstead, 2000; Gross & John, 1998); they are more persuaded by messages appealing to emotion (Meyer & Tormala, 2010); they score higher on measures of empathic concern (for a review, see Eisenberg & Lennon, 1983); and they are more adept at identifying with other's emotional states (e.g., Bullis & Horn, 1995; for a review, see Hall & Schmid Mast, 2008). Together with Greene's dual-process model, these findings suggest that men and women may not differ in terms of their cognitive evaluations of outcomes and thus show equal levels of utilitarian judgments. However, women may experience stronger

affective responses to harm than men, leading to systematic gender differences in deontological judgments.

Despite these links between Greene's (2007) dual-process model and gender differences in affective processing, moral dilemma researchers have largely ignored the role of gender. Most moral dilemma studies do not test gender effects (e.g., Bartels, 2008; Greene et al., 2001). Some studies treated gender as a control variable (e.g., Bartels & Pizarro, 2011), incidentally noting that men showed a stronger preference for utilitarian over deontological judgments than women. We are aware of only one study that directly examined the role of gender in moral dilemma judgments. Fumagalli and colleagues (2010) found that men showed a stronger preference for utilitarian over deontological judgments than women, particularly on "personal" moral dilemmas where harm requires physical force (Greene et al., 2009). No gender differences emerged on non-moral dilemmas or on "impersonal" dilemmas where harm is mediated through mechanical devices (e.g., pressing a button). These findings are consistent with the hypothesis of affect-driven gender differences in moral judgment. However, they can also be explained by gender differences in cognitive evaluations of outcomes, because differences in relative preferences for utilitarian over deontological judgments may reflect differences in utilitarian inclinations, differences in deontological inclinations, or a combination thereof.

PD of Moral Judgments

The ambiguity regarding whether men are more utilitarian than women or women are more deontological than men reflects a confound called the *non-independence error* (Conway & Gawronski, 2013b). According to dual-process models (e.g., Greene et al., 2001), deontological and utilitarian judgments stem from two independent psychological processes: one relatively affective and the other relatively cognitive. Yet, in terms of measurement, utilitarian judgments are presumed to reflect strong utilitarian inclinations and deontological judgments are presumed to reflect strong deontological inclinations. This operationalization treats the two moral inclinations as opposite ends of a bipolar continuum, implying a negative relationship between them. By definition, the more utilitarian judgments a person makes, the fewer deontological judgments he or she makes. Conversely, the more deontological judgments a person makes, the fewer utilitarian judgments he or she makes. Thus, although theorists describe the processes underlying utilitarian and deontological judgments as independent, their measurement is not. To empirically distinguish whether men are more utilitarian than women, or women are more deontological than men, it is necessary to measure deontological and utilitarian inclinations independently.

To resolve the non-independence error, Conway and Gawronski (2013b) adapted Jacoby's (1991) PD procedure to independently estimate the strength of utilitarian and deontological inclinations underpinning overt moral dilemma judgments. Originally developed to examine memory, PD is a content-agnostic procedure that can be applied to any domain where traditional methods conflate the measurement of two psychological processes (for a review, see Payne & Bishara, 2009). Examples of previous applications include the study of racial bias in weapon identification (Payne, 2001), heuristic and rule-based judgments (Ferreira, Garcia-Marques, Sherman, & Sherman, 2006), and the roles of conscious and unconscious thought in decision making (Damian & Sherman, 2013).

A basic requirement for the application of PD is the use of both *incongruent* trials where the two underlying processes lead to divergent responses, as well as *congruent* trials where they lead to the same response. Applied to moral judgments, incongruent dilemmas correspond to high-conflict dilemmas (Koenigs et al., 2007), where action causes immediate harm but leads to a better overall outcome. The crying baby dilemma presented above represents an example of an incongruent dilemma, because acting will kill the baby, thereby saving the rest of the townsfolk. Congruent dilemmas consist of identical structure and content to incongruent dilemmas, except that actions causing immediate harm lead to sub-optimal outcomes overall. For example, the congruent variant of the crying baby dilemma asks whether it is acceptable to smother a baby to prevent townsfolk from laboring in a mine. As killing the baby no longer saves lives, and therefore no longer maximizes overall outcomes, choosing to reject killing the baby is consistent with both deontology and utilitarianism. Thus, in contrast to incongruent dilemmas, where utilitarianism leads to judgments that causing harm is acceptable and deontology leads to judgments that causing harm is unacceptable, on congruent dilemmas both utilitarianism and deontology lead to judgments that causing harm is unacceptable.

Most moral dilemma studies use only incongruent, but not congruent, dilemmas, thereby limiting conclusions to relative preferences for utilitarian over deontological judgments. Yet, the simultaneous use of congruent and incongruent dilemmas makes it possible to disentangle the unique contributions of deontological and utilitarian inclinations to moral judgments. Based on participants' pattern of responses across the two dilemma types, PD algebraically derives two parameters reflecting the strength of deontological and utilitarian inclinations, respectively (see the appendix). It is then possible to investigate the empirical relation between the two inclinations, the relation between each inclination and other variables, and the impact of experimental manipulations on each inclination. For example, Conway and Gawronski (2013b) found that deontological inclinations were uniquely related to individual differences in empathic concern and perspective-taking, whereas utilitarian inclinations were uniquely related to need for cognition. Individual differences in moral identity were positively related to both inclinations, a pattern that was concealed in the traditional approach due to the treatment of the two inclinations as opposite ends of a

bipolar continuum. Moreover, a cognitive load manipulation uniquely reduced utilitarian inclinations, whereas a manipulation enhancing the salience of harm uniquely increased deontological inclinations. These findings corroborate Greene's (2007) dual-process model of moral judgment and demonstrate the utility of PD for tapping deontological and utilitarian inclinations that are conflated in traditional analyses. This unique advantage of PD also helps to resolve the ambiguity underlying gender differences in moral dilemma judgments.

The Current Work

In the current work, we applied PD to clarify whether gender differences in relative preferences for utilitarian over deontological judgments (i.e., relative dilemma judgments) reflect differences in deontological inclinations, differences in utilitarian inclinations, or both. To reach the strongest possible conclusion on this matter, we conducted a meta-analytic re-analysis of 40 available studies, published or unpublished, that used both congruent and incongruent versions of moral dilemmas. Unlike previous work, the use of congruent and incongruent dilemmas allowed us to conduct PD analyses to determine whether (a) men experience stronger utilitarian inclinations than women, (b) women experience stronger deontological inclinations than men, or (c) men and women differ in terms of both moral inclinations. Based on earlier evidence for gender differences in affective processing, we hypothesized that women show stronger deontological inclinations than men. In contrast, we expected no gender differences for utilitarian inclinations, consistent with the lack of gender differences in cognitive processing.

To further inform the interpretation of our results, we also examined gender differences in perceptions of dilemma difficulty. Greene's (2007) dual-process theory implies that incongruent dilemmas should be perceived as more difficult to the extent that a person experiences strong moral inclinations of both kinds. In contrast, incongruent dilemmas should be perceived as less difficult when a person experiences strong inclinations of one kind and weak inclinations of the other kind. Thus, to the extent that women experience both strong deontological and strong utilitarian inclinations, whereas men are guided primarily by utilitarian, but not deontological, inclinations, women should perceive incongruent dilemmas as more difficult than men do.

We also coded each study in our re-analysis according to two sample taxonomies and examined whether these differences moderated the results. First, because college samples tend to be younger and more homogeneous compared with Internet samples, we coded whether each study was conducted with college students or Internet samples. Second, we coded whether each dataset came from our own lab or from other research groups. As we expected findings to be robust across samples and lab groups, we did not expect either taxonomy to moderate the results.

In addition to clarifying the nature of gender differences in moral dilemma judgments, we also aimed to replicate Conway and Gawronski's (2013b) findings that (a) the utilitarian parameter was positively related to relative preferences for utilitarian over deontological judgments, (b) the deontology parameter was negatively related to relative preferences for utilitarian over deontological judgments, and (c) the two parameters were only mildly correlated. If deontological and utilitarian judgments are indeed the result of two independent processes, the parameters representing the two moral inclinations should be un-correlated. Yet, the two parameters should be negatively correlated if moral dilemma responses reflect the opposite end points of a bipolar psychological continuum, as implied by the traditional dilemma approach. In the latter case, the hypothesis of two independent processes would seem questionable, allowing for the possibility that a single-process model provides a more parsimonious theoretical account.

Method

Selection of Studies

The current work entails a meta-analytic re-analysis of existing datasets. These datasets derive primarily from our own lab, but to be as comprehensive as possible, we used three methods to identify other studies for inclusion. First, we searched PsycINFO and Google Scholar using the following parameters: *process dissociation* and *moral** or *dilemma**. Second, we requested relevant data via the email listserv of the Society for Personality and Social Psychology. Third, we contacted relevant authors directly to ask for other published or unpublished datasets. We included all studies containing assessments of (a) moral judgments for both incongruent and congruent dilemmas and (b) participant gender.

Our sample consisted of 40 datasets, including 30 from our lab and 10 from other labs, comprising the total body of results available at that time. These samples contained a total of 6,159 participants. Fifty-nine participants (<1%) failed to indicate gender or did not respond to all dilemmas, leaving us with a final sample of 6,100 participants (see Table 1). Sample sizes ranged from 30 to 397. Ethnicity varied widely across samples. Overall, more than half of the participants identified as Caucasian, approximately one fifth identified as Asian, and relatively few participants identified as Other, Black, East Indian, or Aboriginal.

Materials and Measures

Study authors supplied us with raw dilemma response data, and we subsequently calculated all measurement scores. All studies included participants' responses to the 10 moral dilemmas, each with one congruent and one incongruent version, developed by Conway and Gawronski (2013b). Each dilemma presents an actor who may cause harm to achieve a

Table 1. Overview of Sample Type, Study Design, Sample Size, Age, and Proportions of Gender and Ethnicity in Each Sample.

				Age		Gender %		Ethnicity %					
Sample code number	Study	Sample type	*n*	*M*	*SD*	Women	Men	Caucasian	Asian	Black	East Indian	Aboriginal	Other or unspecified
1	Conway and Gawronski (2013b), Study 1	College	112	19.23	5.20	73.2	26.8	53.6	29.5	2.7	0.0	2.7	11.7
2	Conway and Gawronski (2013b), Study 2	College	57	18.37	0.96	50.9	49.1	56.1	28.1	1.8	12.3	0.0	1.8
3	Conway and Gawronski (2013b), Study 3	Internet	274	34.08	11.73	56.9	43.1	82.5	5.1	5.8	0.4	0.4	5.8
4	Conway and Gawronski (2013d), Study 1	College	73	18.59	0.94	64.4	35.6	67.1	20.5	1.4	2.7	0.0	8.2
5	Conway and Gawronski (2013d), Study 2	College	99	24.31	7.28	68.7	31.3	54.5	27.3	4.0	6.1	0.0	8.0
6	Conway and Gawronski (2013d), Study 3	College	191	19.57	3.99	74.3	25.7	61.8	23.1	4.7	4.2	1.0	5.2
7	Conway and Gawronski (2013c), Study 1	College	74	19.04	2.10	70.3	29.7	79.9	4.1	2.7	0.0	0.0	13.5
8	Conway and Gawronski (2013c), Study 2	College	55	21.58	5.24	74.5	25.5	56.4	25.5	5.5	7.3	0.0	5.5
9	Conway and Gawronski (2013c), Study 3	Internet	353	32.34	10.80	48.4	51.6	50.9	34.4	4.5	4.8	0.3	5.1
10	Conway and Gawronski (2013e), Study 1	College	94	18.71	3.36	73.4	26.6	61.7	25.5	1.1	4.3	1.1	6.4
11	Conway and Gawronski (2013d), Study 4	College	98	22.49	5.16	78.6	21.4	32.7	27.6	6.1	2.0	1.0	30.6
12	Conway and Gawronski (2013d), Study 5	College	128	21.84	4.80	76.6	23.4	40.6	45.3	1.6	1.6	0.0	10.9
13	Conway and Gawronski (2013d), Study 6	Internet	276	34.14	12.16	64.5	35.5	81.5	7.2	6.2	0.0	0.0	5.1
14	Conway and Gawronski (2013d), Study 7	College	140	18.37	1.48	67.9	32.1	60.0	22.9	3.6	1.4	1.4	10.7
15	Conway and Gawronski (2013a), Study 1	College	120	23.78	6.45	75.8	24.2	37.5	35.0	5.0	2.5	1.7	18.3
16	Conway and Gawronski (2013a), Study 2	College	119	18.64	1.54	58.8	41.2	54.6	34.4	0.8	0.0	0.0	10.1
17	Conway, Bartels, and Pizarro (2013), Study 1	Internet	89	34.17	13.06	44.9	55.1	76.1	11.4	3.4	0.0	2.3	6.8
18	Conway et al. (2013), Study 2	Internet	90	36.74	12.30	60.0	40.0	78.4	4.5	11.4	0.0	0.0	5.7
19	Wisneski, Conway, and Skitka (2013), Study 1	Internet	125	36.79	13.27	52.0	48.0	77.6	10.4	2.4	0.8	0.8	8.0
20	Conway, Love, and Mottner (2013), Pre-test	Internet	383	NA	NA	47.8	52.2	NA	NA	NA	NA	NA	NA

(continued)

Table 1. (continued)

Sample code number	Study	Sample type	*n*	Age		Gender %		Ethnicity %					
				M	*SD*	Women	Men	Caucasian	Asian	Black	East Indian	Aboriginal	Other or unspecified
21	Conway et al. (2013), Study 1	Internet	300	NA	NA	47.0	53.0	NA	NA	NA	NA	NA	NA
22	Conway et al. (2013), Study 1	Internet	310	NA	NA	48.1	51.9	NA	NA	NA	NA	NA	NA
23	Friesdorf and Conway (2013), Study 1	Internet	78	29.30	14.71	71.8	28.2	72.7	18.2	6.5	0.0	0.0	2.6
24	Friesdorf and Conway (2013), Study 2	College and Internet	108	26.52	13.35	68.5	31.5	64.5	22.4	9.3	0.0	0.0	3.8
25	Armstrong, Conway, and Gawronski (2013)	Internet	257	35.67	12.14	52.1	47.9	75.0	9.4	6.6	5.5	3.1	0.4
26	Conway and Gawronski (2013a), Study 3	College	240	21.68	6.67	60.0	40.0	41.7	33.3	3.3	6.7	0.8	5.0
27	Peysakhovich, Conway, Rand, and Greene (2013), Study 1	Internet	290	30.72	9.51	51.0	49.0	NA	NA	NA	NA	NA	NA
28	Peysakhovich et al. (2013), Study 2	Internet	397	33.49	12.25	49.9	50.1	NA	NA	NA	NA	NA	NA
29	Robinson, Conway, Gawronski, and Inzlicht (2013)	College	30	18.83	1.12	80.0	20.0	13.3	66.6	6.7	0.0	0.0	13.4
30	Kappes and Van Bavel (2013)	Internet	242	30.81	10.42	40.1	59.9	74.2	8.3	9.1	0.0	0.0	8.4
31	Gamez-Djokic and Molden (2013), Study 1	Internet	73	33.52	12.07	63.0	37.0	58.6	13.3	18.6	0.0	0.0	9.5
32	Gamez-Djokic and Molden (2013), Study 2	Internet	73	33.25	12.41	46.6	53.4	28.7	4.1	56.2	0.0	0.0	11.0
33	Gamez-Djokic and Molden (2013), Study 3	Internet	56	34.95	13.96	44.6	55.4	37.5	1.8	55.4	0.0	0.0	5.3
34	Gamez-Djokic and Molden (2013), Study 4	Internet	107	36.53	14.08	51.4	48.6	NA	NA	NA	NA	NA	NA
35	Gamez-Djokic and Molden (2013), Study 5	College	61	18.75	0.77	50.8	49.2	58.0	6.4	17.5	0.0	0.0	18.1
36	Gamez-Djokic and Molden (2013), Study 6	College	53	18.72	0.89	47.2	52.8	66.7	1.9	24.1	0.0	0.0	7.3
37	Gamez-Djokic and Molden (2013), Study 7	College	56	NA	NA	62.5	37.5	57.1	5.4	25.0	0.0	0.0	12.5
38	Gamez-Djokic and Molden (2013), Study 8	Internet	164	36.32	40.41	54.3	45.7	77.4	7.3	4.9	0.0	0.0	10.4
39	Burgmer, Forstmann, Todd, and Mussweiler (2015)	Internet	87	34.37	13.20	56.3	43.7	NA	NA	NA	NA	NA	NA
40	Forstmann and Conway (2013)	Internet	168	33.08	11.97	37.5	62.5	NA	NA	NA	NA	NA	NA
Total			6,100	27.20	8.93	56.1	43.9	59.03	19.38	9.93	1.96	0.52	9.18

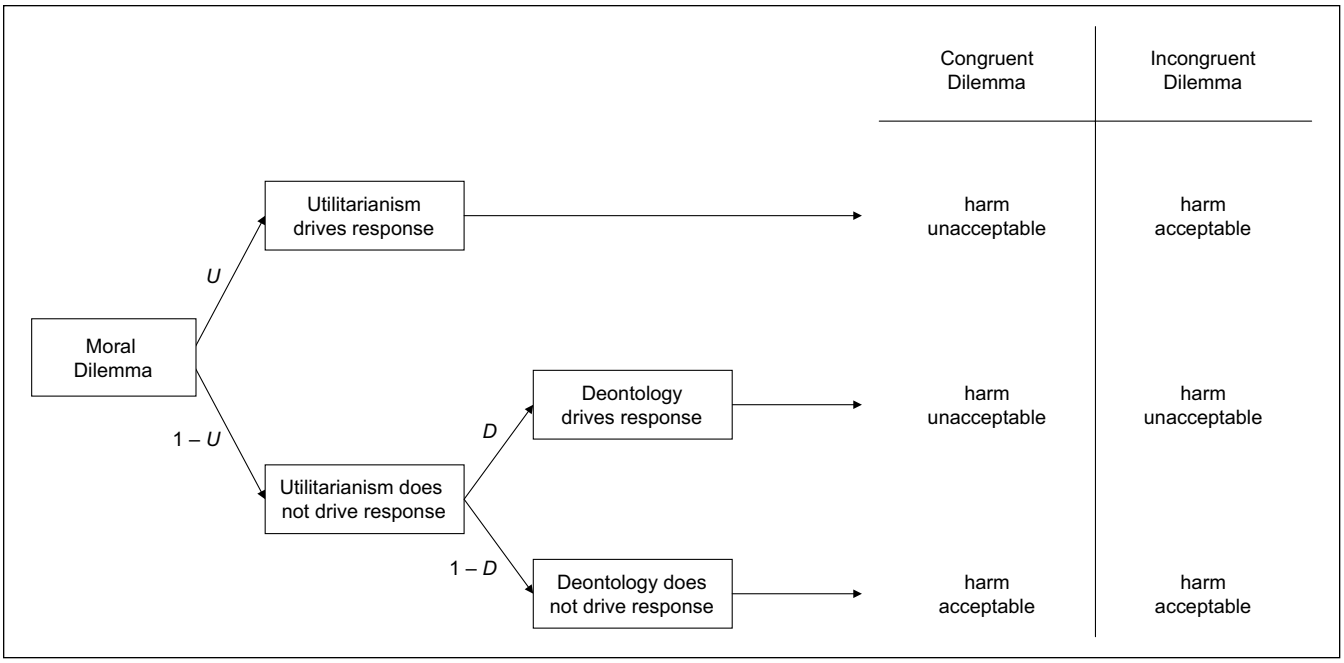

Figure 1. Processing tree illustrating the underlying components leading to judgments that harmful action is either acceptable or unacceptable in congruent and incongruent moral dilemmas.

Source. Copyright 2013 by the American Psychological Association (APA). Reproduced with permission. The official citation that should be used in referencing this material is Conway and Gawronski (2013b). The use of APA information does not imply endorsement by APA.

Note. The paths from left to right depict the three cases that (a) utilitarianism drives the response, (b) deontology drives the response, and (c) neither utilitarianism nor deontology drives the response. The columns on the right depict the potential cases that lead to harm acceptance and harm rejection on congruent and incongruent dilemmas, respectively.

particular outcome: for incongruent dilemmas, action leads to improved overall outcomes, but for congruent dilemmas, action does not. For each dilemma, participants indicated whether the described action was *appropriate* or *not appropriate* (Greene et al., 2001). Traditional relative moral dilemma judgments correspond to the proportion of times participants indicated that harmful action was acceptable for incongruent dilemmas (e.g., Koenigs et al., 2007). In 18 studies, participants additionally indicated how difficult they found reaching a decision for each dilemma, as assessed via a 5-point rating scale ranging from 1 = *very easy* to 5 = *very difficult* (Conway & Gawronski, 2013b).

PD uses answers to congruent as well as incongruent dilemmas. Participants' judgments across both dilemma types can be illustrated via a processing tree (see Figure 1). Utilitarianism entails acting in ways that maximize overall outcomes, whereas deontology entails the rejection of harmful action regardless of outcomes (Foot, 1967). Harmful action maximizes overall outcomes for incongruent, but not congruent, dilemmas. Therefore, utilitarianism and deontology each lead to different patterns of responses across the two dilemma types. For congruent dilemmas, harm will be judged as unacceptable either when utilitarianism drives the response or when deontology drives the response. Conversely, harm will be judged as acceptable on congruent dilemmas when neither utilitarianism nor deontology drive the response. For incongruent dilemmas, harm will be judged as unacceptable when deontology drives the response. Conversely, harm will be judged acceptable either when utilitarianism drives the response, or when neither utilitarianism nor deontology drives the response. Conceptualizing dilemma responses in this manner allows researchers to represent the probability that participants indicated harm was acceptable in congruent and incongruent dilemmas, and combine the resulting equations to algebraically solve for the parameters U and D (see the appendix).[1]

Statistical Analysis

We began by computing the proportion of times women and men deemed causing harm acceptable on incongruent dilemmas, in line with traditional scoring of moral dilemma data (e.g., Koenigs et al., 2007). This proportion can be interpreted as a measure of participants' relative preference for utilitarian over deontological judgments when the two moral principles are in conflict. We computed the difference between women's and men's judgments in terms of Cohen's d using pooled standard deviations for each dataset. Next, we estimated the deontological and utilitarian PD parameters and computed the Cohen's d for gender differences on each parameter. All mean difference comparisons were corrected for small sample size bias (Lipsey & Wilson, 2001).

Then we computed the meta-analytic effect size of gender differences on each variable via the SPSS *MEAN ES* macro (Wilson, 2005). We report confidence intervals (CIs) based on the random-effects method of moments estimation models for all analyses to draw inferences beyond the existing sample of studies. To account for variations in sample size, we weighted each effect size by the inverse of its variance, which is a more accurate index of effect size precision than is sample size (Marín-Martínez & Sánchez-Meca, 2010). For consistency, we always subtracted men's scores from women's scores. Thus, negative values indicate that men showed higher scores than women on a given measure; positive values indicate that women showed higher scores than men on a given measure.

In 18 of the 40 studies included in the current analysis, participants also indicated how difficult it was for them to reach a decision for each dilemma. To further constrain the interpretation of our findings, we aggregated these ratings separately for men and women and computed Cohen's *d* to examine gender differences in difficulty perceptions.

In addition to our main analyses, we computed the correlations between relative preferences for utilitarian over deontological judgments, utilitarian inclinations, and deontological inclinations in each study. Correlations were *Z*-transformed for analysis and weighted by the inverse variance for correlations when comparing three pairs of scores: $n - 3$ (Lipsey & Wilson, 2001).

We examined heterogeneity of variance for each of the effect sizes under consideration. In cases where there was substantial heterogeneity of variance, we tested whether the reported effects were moderated by sample type (Internet vs. college; own lab vs. other labs). Finally, we examined whether the effect sizes under consideration were correlated with the sample size and gender ratio of each study to determine whether these factors affect results.

Results

Moral Judgments

Relative preferences. The means and standard deviations of relative preferences for utilitarian over deontological judgments are presented separately for men and women in Table 2, including the effect sizes of gender differences in each study. We expected to replicate previous findings that men show a stronger preference for utilitarian over deontological judgments than women. Indeed, the Cohen's *d* was always negative, ranging from −0.94 to −0.12 (weighted *SD* = 0.16), indicating that, as predicted, men consistently showed a stronger preference for utilitarian over deontological judgments than did women. The summary statistics confirmed this finding. The meta-analytic Cohen's *d* across studies was −0.52 (*SE* = .03). This effect size qualifies as moderate according to Cohen's (1988) guidelines. The 95% CI excluded zero [−.57, −.47] and a *z*-test further confirmed that the effect was significantly different from zero, $z = -19.35$, $p < .001$. Cochran's homogeneity test was not significant, $Q(39) = 34.95$, $p = .655$, indicating that differences between the datasets did not contribute to significant variation in effect sizes. For a more precise estimate, we calculated the I^2 statistic, which estimates the amount of variance in effect sizes attributable to systematic differences across samples. As this statistic was negative, it was rescored as zero, suggesting that little of the variance in effect sizes was attributable to systematic differences across datasets (Higgins, Thompson, Deeks, & Altman, 2003).

Utilitarian inclinations. As discussed above, gender differences in the preference for utilitarian over deontological judgments could be driven by gender differences in utilitarian inclinations, gender differences in deontological inclinations, or both. However, based on small gender differences in cognitive processing, more generally, we anticipated only small gender differences on the utilitarian parameter. The Cohen's *d* for utilitarian inclinations varied considerably, ranging from −0.90 to 0.43 (weighted *SD* = 0.21), indicating that in some samples men showed stronger utilitarian inclinations than women, whereas in other samples the opposite obtained (see Table 2). As predicted, across samples, men scored slightly higher in utilitarian inclinations than women, with a meta-analytic Cohen's *d* of −0.10 (*SE* = .03). This effect size falls below Cohen's benchmark for a small effect. Nonetheless, this small effect was significantly different from zero, $z = -2.83$, $p = .005$, and the 95% CI excluded zero [−.17, −.03]. Cochran's homogeneity test was significant, $Q(39) = 62.73$, $p = .009$, and the I^2 statistic indicated that approximately 38% of the variance in effect sizes was due to systematic variation in the samples, $I^2 = 37.82$, which is between a small and moderate amount (Higgins et al., 2003). Therefore, we conducted further analyses to test whether utilitarian inclinations were moderated by sample type (see below).

Deontological inclinations. In contrast to utilitarian inclinations, we expected women to score substantially higher than men on deontological inclinations, based on gender differences in affective processing more generally. As predicted, women consistently showed stronger deontological inclinations than men (see Table 2), with Cohen's *d*s ranging from 0.12 to 0.85 (weighted *SD* = 0.15). The meta-analytic Cohen's *d* was 0.57 (*SE* = .03). This effect qualifies as slightly larger than moderate, according to Cohen's guidelines. The effect was significantly different from zero, $z = 21.15$, $p < .001$, and the 95% CI excluded zero [.52, .62]. Cochran's homogeneity test was not significant, $Q(39) = 31.67$, $p = .791$, and I^2 was negative (thus set to zero), suggesting that differences between the datasets did not contribute to significant variation in the effect sizes.

Difficulty Ratings

Table 3 shows the means and standard deviations of difficulty ratings for the 18 samples where such ratings were collected ($n = 1{,}837$). We hypothesized that women would

Table 2. Means, Standard Deviations, and Effect Sizes of Gender Differences for Relative Preference for Utilitarian Over Deontological Judgments, Utilitarian Inclinations, and Deontological Inclinations.

			Preference for utilitarian over deontological judgments			Utilitarian inclinations (standardized)			Deontological inclinations (standardized)		
Sample code number	Women, *n*	Men, *n*	Women	Men	Cohen's *d*	Women	Men	Cohen's *d*	Women	Men	Cohen's *d*
1	82	30	.55 (0.18)	.65 (0.17)	−0.52	.05 (0.88)	−.15 (1.28)	0.20	.20 (0.94)	−.54 (0.97)	0.77
2	29	28	.56 (0.15)	.67 (0.13)	−0.73	−.13 (1.00)	.14 (1.00)	−0.27	.25 (0.99)	−.26 (0.96)	0.52
3	156	118	.51 (0.18)	.62 (0.17)	−0.64	−.10 (1.02)	.13 (0.97)	−0.23	.24 (0.81)	−.32 (1.13)	0.58
4	47	26	.56 (0.14)	.69 (0.14)	−0.95	−.18 (0.95)	.33 (1.03)	−0.53	.24 (0.88)	−.43 (1.07)	0.70
5	68	31	.55 (0.14)	.66 (0.15)	−0.82	−.21 (0.92)	.47 (1.03)	−0.71	.12 (1.00)	−.27 (0.95)	0.39
6	142	49	.57 (0.14)	.60 (0.20)	−0.19	−.01 (0.97)	.03 (1.10)	−0.04	.08 (0.92)	−.24 (1.19)	0.33
7	52	22	.55 (0.17)	.60 (0.17)	−0.33	−.02 (0.90)	.06 (1.22)	−0.08	.12 (1.05)	−.28 (0.84)	0.39
8	41	14	.55 (0.15)	.64 (0.12)	−0.65	−.01 (0.99)	.02 (1.06)	−0.03	.20 (0.95)	−.58 (0.95)	0.81
9	171	182	.51 (0.16)	.62 (0.17)	−0.70	.08 (0.88)	−.06 (1.07)	0.14	.41 (0.82)	−.36 (0.99)	0.85
10	69	25	.56 (0.15)	.60 (0.15)	−0.29	.01 (0.99)	−.01 (1.05)	0.01	.09 (0.95)	−.26 (1.11)	0.35
11	77	21	.56 (0.14)	.64 (0.12)	−0.59	−.07 (1.00)	.26 (1.00)	−0.33	.05 (0.94)	−.20 (1.21)	0.25
12	98	30	.57 (0.15)	.62 (0.18)	−0.34	.03 (0.99)	−.08 (1.03)	0.11	.16 (0.96)	−.51 (0.98)	.69
13	178	98	.55 (0.17)	.64 (0.16)	−0.58	−.08 (1.00)	.15 (0.98)	−0.23	.21 (0.97)	−.39 (0.95)	0.63
14	95	45	.56 (0.15)	.63 (0.14)	−0.44	.01 (0.93)	−.02 (1.15)	0.02	.17 (0.96)	−.35 (1.00)	0.59
15	91	29	.56 (0.15)	.65 (0.16)	−0.60	−.01 (0.97)	.03 (1.11)	−0.04	.16 (0.87)	−.52 (1.21)	0.71
16	70	49	.57 (0.15)	.62 (0.15)	−0.33	.12 (0.97)	−.17 (1.03)	0.29	.24 (0.90)	−.35 (1.04)	0.61
17	40	49	.54 (0.18)	.63 (0.18)	−0.54	−.02 (0.77)	.01 (1.17)	−0.03	.31 (0.93)	−.26 (0.99)	0.59
18	54	36	.49 (0.22)	.67 (0.16)	−0.90	−.18 (0.94)	.27 (1.04)	−0.46	.31 (0.96)	−.46 (0.89)	0.82
19	65	60	.51 (0.18)	.62 (0.21)	−0.56	−.03 (0.96)	.03 (1.05)	−0.07	.29 (0.89)	−.32 (1.02)	0.63
20	183	200	.55 (0.17)	.62 (0.17)	−0.43	−.01 (0.98)	.01 (1.02)	−0.03	.28 (0.93)	−.26 (1.00)	0.56
21	141	159	.52 (0.19)	.63 (0.17)	−0.57	−.09 (0.99)	.08 (1.00)	−0.17	.29 (0.89)	−.26 (1.03)	0.57
22	149	161	.55 (0.18)	.63 (0.17)	−0.50	−.11 (0.94)	.10 (1.05)	−0.21	.22 (0.96)	−.21 (0.99)	0.44
23	56	22	.51 (0.19)	.56 (0.20)	−0.27	.03 (0.91)	−.07 (1.23)	0.09	.13 (1.01)	−.34 (0.92)	0.48
24	74	34	.53 (0.19)	.60 (0.19)	−0.39	−.05 (0.95)	.10 (1.11)	−0.15	.09 (1.00)	−.21 (0.99)	0.30
25	134	123	.52 (0.19)	.64 (0.18)	−0.63	−.06 (0.98)	.07 (1.02)	−0.13	.35 (0.89)	−.38 (0.97)	0.79
26	144	96	.56 (0.16)	.60 (0.19)	−0.24	−.01 (0.94)	.01 (1.09)	−0.01	.14 (0.96)	−.22 (1.03)	0.37
27	148	142	.51 (0.20)	.61 (0.20)	−0.48	−.04 (0.94)	.05 (1.06)	−0.09	.26 (0.88)	−.27 (1.05)	0.55
28	198	199	.52 (0.18)	.64 (0.17)	−0.65	−.06 (0.94)	.06 (1.06)	−0.13	.34 (0.94)	−.34 (0.95)	0.72
29	24	6	.53 (0.22)	.68 (0.12)	−0.75	−.17 (0.99)	.69 (0.78)	−0.91	.09 (1.07)	−.34 (0.63)	0.42
30	97	145	.56 (0.18)	.64 (0.17)	−0.46	−.03 (0.93)	.02 (1.05)	−0.04	.29 (0.95)	−.20 (0.99)	0.50
31	46	27	.54 (0.17)	.62 (0.20)	−0.48	−.04 (1.02)	.06 (0.98)	−0.10	.20 (0.86)	−.34 (1.14)	0.56
32	34	39	.52 (0.16)	.56 (0.25)	−0.19	−.05 (0.89)	.04 (1.10)	−0.09	.15 (0.87)	−.13 (1.09)	0.28
33	25	31	.55 (0.13)	.61 (0.19)	−0.36	.23 (0.87)	−.19 (1.07)	0.44	.38 (0.78)	−.30 (1.06)	0.71
34	55	52	.57 (0.16)	.65 (0.17)	−0.48	−.05 (0.84)	.05 (1.14)	−0.10	.23 (0.84)	−.24 (1.10)	0.49
35	31	30	.54 (0.17)	.64 (0.15)	−0.63	−.14 (0.91)	.15 (1.08)	−0.29	.26 (0.92)	−.26 (1.02)	0.54
36	25	28	.62 (0.17)	.64 (0.14)	−0.12	.01 (1.09)	−.01 (0.93)	0.01	.06 (0.97)	−.05 (1.04)	0.12
37	35	21	.58 (0.22)	.67 (0.15)	−0.47	−.01 (1.05)	.02 (0.94)	−0.03	.22 (1.07)	−.37 (0.76)	0.62
38	89	75	.56 (0.17)	.64 (0.14)	−0.55	−.02 (0.93)	.02 (1.08)	−0.04	.21 (0.96)	−.25 (1.00)	0.47
39	49	38	.54 (0.17)	.67 (0.14)	−0.77	−.25 (1.00)	.32 (0.91)	−0.59	.22 (0.98)	−.28 (0.96)	0.51
40	63	105	.51 (0.21)	.62 (0.20)	−0.53	−.09 (0.96)	.05 (1.02)	−0.14	.28 (0.87)	−.16 (1.03)	0.46
Total	3,425	2,675			−0.52			−0.10			0.57

Note. Standard deviations are presented in parentheses.

perceive incongruent dilemmas as more difficult than men do, to the extent that women experience both strong deontological and strong utilitarian inclinations more so than men. The data confirmed this hypothesis: Women rated incongruent dilemmas as more difficult than did men. Cohen's *d* varied between −0.09 and 0.83 (weighted *SD* = 0.23), and the meta-analytic effect size was .38 (*SE* = .05), a small effect according to Cohen. Yet, the effect was significantly different from zero, $z = 6.95$, $p < .001$, and the 95% CI excluded zero [.28, .49]. Cochran's homogeneity test was not significant, $Q(17) = 19.19$, $p = .318$, and $I^2 = 11.41$.

Although we expected that women would rate incongruent dilemmas as more difficult than would men, we had no specific hypothesis regarding gender differences in perceptions of congruent dilemma difficulty. Our analyses revealed t hat women also rated congruent dilemmas as more difficult than did men. The Cohen's *d* for women's versus men's difficulty ratings varied between −0.02 and 0.75 (weighted *SD* = 0.17).

Table 3. Means, Standard Deviations, and Effect Sizes of Gender Differences for Perceived Decision Difficulty on Congruent and Incongruent Dilemmas.

	Incongruent dilemmas			Congruent dilemmas		
Sample code number	Women	Men	Cohen's *d*	Women	Men	Cohen's *d*
1	2.97 (0.62)	2.68 (0.64)	0.47	2.51 (0.58)	2.50 (0.58)	0.01
2	3.24 (0.47)	2.85 (0.47)	0.83	2.54 (0.49)	2.42 (0.44)	0.27
4	3.21 (0.51)	2.91 (0.64)	0.53	2.65 (0.59)	2.49 (0.66)	0.26
5	2.92 (0.65)	2.93 (0.52)	−0.01	2.51 (0.60)	2.34 (0.50)	0.31
6	3.03 (0.60)	2.90 (0.73)	0.21	2.52 (0.61)	2.51 (0.73)	0.02
7	3.17 (0.50)	2.72 (0.61)	0.84	2.52 (0.49)	2.31 (0.52)	0.42
8	3.02 (0.61)	2.59 (0.61)	0.71	2.46 (0.52)	2.39 (0.57)	0.13
10	3.13 (0.61)	2.80 (0.57)	0.56	2.55 (0.62)	2.31 (0.47)	0.41
11	3.01 (0.59)	3.06 (0.66)	−0.09	2.54 (0.67)	2.55 (0.77)	−0.02
12	2.97 (0.57)	2.68 (0.69)	0.48	2.50 (0.58)	2.34 (0.76)	0.25
14	3.07 (0.61)	2.88 (0.63)	0.31	2.54 (0.60)	2.56 (0.63)	−0.02
15	3.00 (0.61)	2.80 (0.95)	0.28	2.46 (0.59)	2.47 (0.85)	−0.01
16	2.95 (0.62)	2.62 (0.63)	0.54	2.45 (0.63)	2.32 (0.57)	0.21
26	2.94 (0.48)	2.79 (0.57)	0.29	2.36 (0.59)	2.35 (0.49)	0.01
29	2.88 (0.69)	2.42 (0.71)	0.66	2.63 (0.64)	2.15 (0.48)	0.78
31	2.89 (0.71)	2.79 (0.59)	0.15	2.45 (0.51)	2.61 (0.56)	0.30
32	3.03 (0.51)	2.65 (0.75)	0.59	2.67 (0.74)	2.68 (0.75)	−0.02
35	3.49 (0.36)	3.24 (0.65)	0.47	2.82 (0.54)	2.58 (0.60)	0.42
Total			0.38			0.15

Note. Standard deviations are presented in parentheses. Missing sample code numbers indicate samples that did not contain difficulty rating data.

The meta-analytic effect size was .15 (SE = .05), which falls below the benchmark of a small effect according to Cohen (1988). Nonetheless, the effect was significantly different from zero, $z = 2.84$, $p = .004$, and the 95% CI excluded zero [.04, .24]. Cochran's homogeneity test was not significant, $Q(17) = 11.14$, $p = .849$, and I^2 was negative (thus set to zero).

Although women rated both congruent and incongruent dilemmas as more difficult than did men, both women and men rated incongruent dilemmas as more difficult than congruent ones. Moreover, the difference between difficulty ratings of congruent and incongruent dilemmas was larger for women than men, with $d = 1.02$ (weighted $SD = 0.20$) for women and $d = 0.72$ (weighted $SD = 0.34$) for men. The difference in effect sizes was statistically significant, $t(19) = 4.78$, $p < .001$.

Correlations

Table 4 presents the correlations between relative preferences for utilitarian over deontological judgments, the PD parameter for utilitarian inclinations, and the PD parameter for deontological inclinations. We expected that utilitarian inclinations would positively correlate with relative dilemma judgments, and deontological inclinations would negatively correlate with relative dilemma judgments, but the two parameters would be only mildly correlated or un-correlated. All predictions were confirmed. Utilitarian inclinations correlated positively with relative judgments in all studies, with correlations ranging from .43 to .76. The overall meta-analytic correlation was .60, 95% CI = [.58, .63], which qualifies as a large effect according to Cohen, and significantly different from zero, $z = 37.17$, $p < .001$. Cochran's test of heterogeneity was significant, $Q(39) = 71.77$, $p = .001$, and $I^2 = 45.66$, suggesting that a moderate amount of the variance in effect sizes was due to sample differences (Higgins et al., 2003), so we conducted a moderation analysis (see below).

Also as expected, deontological inclinations correlated negatively with relative preferences for utilitarian over deontological judgments in all samples, ranging from −.47 to −.85. The overall meta-analytic correlation was −.71, 95% CI = [−.71, −.68], a large effect (Cohen, 1988) that was significantly different from zero, $z = -36.06$, $p < .001$. Cochran's homogeneity test was significant, $Q(39) = 127.17$, $p < .001$, and $I^2 = 69.33$, indicating that a large amount of variation in effect sizes was due to sample differences, so we conducted a moderation analysis (see below).

Although the PD parameters correlated highly with relative preferences for utilitarian over deontological judgments, the parameters themselves were un-correlated in most samples. Correlations ranged from −.31 to .36, with a meta-analytic correlation of .10, 95% CI = [.07, .13], a small effect (Cohen, 1988) that was nonetheless significantly different from zero, $z = 6.41$, $p < .001$. The heterogeneity test was non-significant, $Q(39) = 53.02$, $p = .067$, and $I^2 = 26.42$, suggesting that sample

Table 4. Correlations Between Relative Preference for Utilitarian Over Deontological Judgments, Utilitarian Inclinations, and Deontological Inclinations.

Sample code number	Relative preference and utilitarian inclinations	Relative preference and deontological inclinations	Utilitarian and deontological inclinations
1	.56***	−.77***	.09
2	.60***	−.67***	.12
3	.69***	−.66***	.04
4	.65***	−.63***	.14
5	.72***	−.60***	.08
6	.67***	−.59***	.14*
7	.65***	−.70***	.05
8	.64***	−.67***	.11
9	.49***	−.67***	.27***
10	.68***	−.48***	.27**
11	.61***	−.47***	.36***
12	.68***	−.63***	.11
13	.66***	−.67***	.09
14	.58***	−.66***	.18*
15	.60***	−.72***	.09
16	.56***	−.71***	.15
17	.49***	−.77***	.12
18	.64***	−.75***	−.01
19	.47***	−.81***	.10
20	.64***	−.73***	.02
21	.63***	−.69***	.09
22	.59***	−.69***	.12
23	.61***	−.67***	.14
24	.60***	−.66***	.16
25	.67***	−.69***	.03
26	.65***	−.66***	.11
27	.61***	−.78***	−.03
28	.51***	−.80***	.05
29	.76***	−.84***	−.31
30	.57***	−.69***	.15*
31	.57***	−.79***	.03
32	.45***	−.85***	.01
33	.43**	−.83***	.12
34	.66***	−.67***	.09
35	.69***	−.74***	−.06
36	.57***	−.77***	.04
37	.56***	−.78***	.01
38	.48***	−.64***	.31***
39	.63***	−.76***	−.01
40	.50***	−.81***	.05
Total	.60***	−.70***	.10***

*$p < .05$. **$p < .01$. ***$p < .001$.

differences did not affect correlation strength. The overall pattern of meta-analytic correlations was similar, albeit slightly stronger, for women than men (see Table 5).

Moderation Analyses

College Versus Internet. Although most effect sizes were homogeneous, gender differences in utilitarian inclinations exhibited substantial heterogeneity of variance across samples. In an attempt to explain some of this variance, we tested whether effect sizes significantly differed for college and Internet samples. We did not expect sample type to moderate the results. The Cohen's *d* for college samples ($n = 19$) varied between −0.71 and 0.44 (weighted $SD = 0.24$), with a meta-analytic effect size of −.12, 95% CI = [−.24, −.01]. The Cohen's *d* for Internet samples ($n = 21$) varied between −0.91

Table 5. Correlations Between Relative Preference for Utilitarian Over Deontological Judgments, Utilitarian Inclinations, and Deontological Inclinations for Women and Men.

	Relative preference and utilitarian inclinations		Relative preference and deontological inclinations		Utilitarian and deontological inclinations	
Sample code number	Women	Men	Women	Men	Women	Men
1	.61***	.61***	−.76***	−.64***	.01	.16
2	.58**	.56**	−.57**	−.51**	.30	.35
3	.83***	.49***	−.59***	−.69***	−.09	.23*
4	.63***	.59**	−.57***	−.66***	.30*	.15
5	.64***	.76***	−.56***	−.64***	.24*	−.04
6	.64***	.72***	−.51***	−.73***	.30***	−.15
7	.59***	.79***	−.72***	−.61**	.10	−.02
8	.70***	.59*	−.67***	−.52*	.04	.37
9	.63***	.46***	−.58***	−.68***	.24**	.29***
10	.72***	.62**	−.44***	−.54**	.28*	.27
11	.62***	.53*	−.59***	−.29	.30**	.63**
12	.68***	.74***	−.55***	−.81***	.21*	−.23
13	.63***	.70***	−.62***	−.68***	.18*	.01
14	.60***	.61***	−.67***	−.58***	.15	.24
15	.65***	.53**	−.70***	−.78***	.11	.08
16	.67***	.49***	−.65***	−.76***	.10	.12
17	.40**	.50***	−.81***	−.73***	.18	.15
18	.63***	.61***	−.73***	−.65***	.03	.15
19	.52***	.44***	−.74***	−.82***	.14	.09
20	.68***	.60***	−.68***	−.75***	.03	.03
21	.70***	.56***	−.66***	−.67***	.04	.18*
22	.60***	.58	−.66***	−.69***	.16	.14
23	.61***	.49*	−.66***	−.75***	.16	.18
24	.61***	.46**	−.67***	−.67***	.14	.29
25	.73***	.58***	−.58***	−.76	.10	.04
26	.62***	.69***	−.62***	−.69***	.19*	.02
27	.74***	.52***	−.75***	−.78***	−.14	.08
28	.51***	.52***	−.77***	−.80***	.13	.04
29	.73***	.77	−.85***	−.74	−.28	−.16
30	.61***	.52***	−.68***	−.69***	.13	.20*
31	.64***	.49**	−.71***	−.84***	.07	.01
32	.28	.53***	−.76***	−.89***	.37*	−.18
33	.53**	.47**	−.72***	−.88***	.20	−.02
34	.71***	.63***	−.72***	−.59***	−.06	.20
35	.69***	.70***	−.80***	−.64***	−.14	.08
36	.69***	.43*	−.73***	−.82***	−.04	.12
37	.57***	.58***	−.76***	−.76***	−.01	.06
38	.57***	.41***	−.69***	−.54***	.16	.49***
39	.54***	.67***	−.77***	−.72***	.08	.04
40	.64***	.42***	−.83***	−.79***	−.14	.16
Total	.64***	.55***	−.66***	−.71***	.04**	.05**

*$p < .05$. **$p < .01$. ***$p < .001$.

and 0.29 (weighted $SD = 0.20$), with a meta-analytic effect size of −0.09, 95% CI = [−.18, −.01]. As predicted, these effect sizes did not significantly differ, $z = 0.43$, $p = .314$. Cochran's homogeneity test remained significant for Internet samples, $Q(20) = 39.61$, $p = .006$, but not for college samples, $Q(18) = 23.71$, $p = .165$.

In addition, two correlations exhibited substantial heterogeneity across samples. We first examined whether the correlation between the utilitarian parameter and relative preferences for utilitarian over deontological judgments varied as a function of sample type. For Internet samples, this correlation was .58, 95% CI = [.56, .61], $z = 29.91$, $p < .001$,

whereas for college samples, it was .63, 95% CI = [.60, .67], $z = 26.09$, $p < .001$. As predicted, these effect sizes did not significantly differ, $z = -0.23$, $p = .818$. Cochran's homogeneity test was no longer significant for either Internet, $Q(20) = 26.22$, $p = .159$, or college samples, $Q(18) = 9.40$, $p = .950$.

Next, we examined whether the correlation between the deontological parameter and relative preferences for utilitarian over deontological judgments varied across sample type. For Internet samples, this correlation was −.73, 95% CI = [−.76, −.71], $z = -31.77$, $p < .001$, whereas for college samples, it was −.67, 95% CI = [−.71, −.63], $z = 23.10$, $p < .001$. Again, as predicted, these effect sizes did not significantly differ, $z = -0.40$, $p = .689$. Cochran's homogeneity test was no longer significant for either Internet, $Q(20) = 22.27$, $p = .326$, or college samples, $Q(18) = 20.10$, $p = .327$.

Research group. Thirty of the included datasets came from our research group, whereas 10 came from other labs. Therefore, we examined whether the source of the dataset moderated the findings that showed systematic variation across samples. We expected that findings would be robust across research groups. For gender differences in utilitarian inclinations, the meta-analytic Cohen's d was −0.10 ($SE = .04$), 95% CI = [−.18, −.02], $z = -2.53$, $p = .011$, for samples from our lab, and −0.11 ($SE = .08$), 95% CI = [−.26, −.05], $z = -1.38$, $p = .165$, for samples from other research groups. As predicted, these effect sizes did not significantly differ, $z = -0.11$, $p = .909$, and Cochran's homogeneity test was not significant for either group, $Q(29) = 27.45$, $p = .440$ and $Q(9) = 12.20$, $p = .202$, respectively.

The correlation between the utilitarian parameter and relative preferences for utilitarian over deontological judgments was .61, 95% CI = [.59, .64], $z = 35.78$, $p < .001$, for samples from our lab, and .57, 95% CI = [.51, .62], $z = 15.76$, $p < .001$, for samples from other research groups. As predicted, these effect sizes did not significantly differ, $z = 0.14$, $p = .889$. Cochran's homogeneity test was no longer significant for either sample, $Q(29) = 28.59$, $p = .464$ and $Q(9) = 8.06$, $p = .528$, respectively.

The correlation between the deontological parameter and relative preferences for utilitarian over deontological judgments was −.70, 95% CI = [−.72, −.67], $z = -31.30$, $p < .001$, for samples from our lab, and −.75, 95% CI = [−.79, −.70], $z = -18.58$, $p < .001$, for samples from other research groups. As predicted, these effect sizes did not significantly differ, $z = 0.25$, $p = .802$. Cochran's homogeneity test was no longer significant for either sample, $Q(29) = 31.43$, $p = .345$ and $Q(9) = 8.77$, $p = .458$, respectively.

Sampling Bias

To investigate whether our findings are distorted by differences in sample sizes and gender ratios, we correlated the effect sizes of gender differences in relative judgments and the two PD parameters with (a) the total sample size and (b) the percentage of women in each study. Although gender differences in relative dilemma judgments showed a marginally significant positive correlation with percentage of women ($r = .30$, $p = .06$), none of the other correlations reached significance (all rs < .26, all ps > .10). These results suggest that sampling bias did not contribute to the findings of our meta-analytic re-analysis, at least for gender differences in the two PD parameters.

Discussion

Our meta-analytic re-analysis of 40 studies containing 6,100 participants indicated that men showed a stronger preference for utilitarian over deontological judgments than women, replicating previous findings (e.g., Bartels & Pizarro, 2011; Fumagalli et al., 2010). Measures of heterogeneity suggested that variation across samples did not contribute appreciably to the size of this effect. Taken at face value, this result may indicate that men engage in more cognitive processing aimed at maximizing overall outcomes than women. Alternatively, the obtained difference might indicate that women experience stronger affective reactions to the idea of causing harm than men. Traditional approaches are unable to resolve this ambiguity. Gender differences in relative preferences could be driven by stronger utilitarian inclinations among men, stronger deontological inclinations among women, or a combination of these effects.

To resolve this ambiguity, we conducted a PD analysis that independently measured deontological and utilitarian inclinations underpinning moral dilemma judgments (Conway & Gawronski, 2013b). Our analysis indicated that women scored substantially higher than men on the parameter reflecting deontological inclinations, whereas men scored only slightly higher than women on the parameter reflecting utilitarian inclinations. This result suggests that most of the variance in relative preferences for utilitarian over deontological judgments stems from gender differences in affective reactions to causing harm; far less variance in relative preferences stems from gender differences in cognitive evaluations of outcomes. In other words, when faced with a moral dilemma, men and women tend to engage in similar degrees of utilitarian processing. Yet, women are more likely to engage in deontological processing than men.

It is worth comparing the size of these effects to those found in other meta-analyses. Hyde (2005) classified the effect sizes of gender differences in 124 meta-analyses in terms of five categories: close-to-zero ($d \leq 0.10$), small ($0.11 < d < 0.35$), moderate ($0.36 < d < 0.65$), large ($0.66 < d < 1.00$), and very large ($d > 1.00$). According to this rubric, gender differences in utilitarian inclinations fall just below the benchmark for the small category. In contrast, gender differences in deontological inclinations fall squarely within the moderate category—which makes them larger than at least 78% of all gender differences reported in Hyde's study. In a much broader review of

meta-analyses in psychology, Richard, Bond, and Stokes-Zoota (2003) noted that meta-analytic effect sizes for gender differences tend to be much smaller (average r = .12 based on 5,691 samples) than meta-analytic effect sizes for other findings in social psychology (average r = .22 based on 28,222 samples). Thus, by converting Cohen's d to r, gender differences on the utilitarian parameter are smaller than the average effect size of gender differences in psychology, r = .05, whereas gender differences on the deontology parameter are larger than the average effect size, r = .27. Thus, gender differences in utilitarianism appear negligible, whereas those on deontology constitute a sizable effect. Notably, gender differences on the deontology parameter are also considerably larger than the gender differences in care and justice orientations found by Jaffee and Hyde (2000), suggesting that previous work on moral reasoning underestimated gender differences in moral judgment.

In addition to moral dilemma judgments, we analyzed participants' perceptions of dilemma difficulty for both congruent and incongruent dilemmas in the 18 studies where such data were available (n = 1,837). If women engage in both utilitarian and deontological processing, they should experience more conflict than men when deontological and utilitarian inclinations imply conflicting decisions. Consistent with this hypothesis, women reported perceiving incongruent dilemma decisions as more difficult than did men, and the difference in perceived difficulty between congruent and incongruent dilemmas was larger for women than men. Together with the obtained gender differences in the two PD parameters, these findings corroborate the conclusion that women experience both strong deontological and strong utilitarian inclinations, whereas men's responses seem to be influenced by strong utilitarian inclinations and weak deontological inclinations.

We also examined the meta-analytic correlations between the two moral inclinations, as well as their correlation with relative preferences for utilitarian over deontological judgments. According to dual-process models (e.g., Greene et al., 2001), deontological and utilitarian judgments stem from two functionally independent processes, one being affective and the other cognitive. Although traditional measurement techniques treat these responses as opposite ends of a bipolar continuum, PD allows for an empirical examination of the proposed independence. Mathematically, the parameters are free to covary, and therefore researchers may examine the degree of correlation between them. If the dual-process model is correct, the parameters ought to be largely un-correlated. In contrast, a single-process model of moral judgment would imply that the parameters should be highly negatively correlated. The results of our meta-analytic re-analysis confirmed that the correlation between utilitarian and deontological inclinations was very small (and slightly positive), corroborating the claim that the two moral inclinations stem from relatively independent processes. Moreover, the utilitarian parameter correlated positively and the deontology parameter correlated negatively with relative preferences for utilitarian over deontological judgments, indicating that the variance captured by each parameter is confounded in the traditional measurement approach. This pattern emerged for both men and women.

Finally, we examined heterogeneity of variance for each effect. In most cases, the Cochran's test for homogeneity of variance was not significant, indicating that variation across the datasets did not significantly affect the size of the reported effects. Moreover, we calculated the I^2 statistic to estimate the proportion of variance in effect sizes attributable to variation across the datasets. This statistic was smaller than 25 for most of the reported effects, indicating that less than 25% of the variance in effect sizes was attributable to variation across the datasets. This number corresponds to a small amount according to the guidelines suggested by Higgins and colleagues (2003). Therefore, most of the reported results were robust with regard to differences across samples. Three exceptions were (a) the amount of variation obtained for gender differences in utilitarian inclinations, for which approximately 39% of the variance in effect sizes was due to systematic variation in the samples; (b) the correlation between the utilitarian parameter and relative judgments; and (c) the correlation between the deontology parameter and relative judgments. In all three cases, the obtained effects remained consistent across the different sub-samples. Moreover, neither of the two PD parameters was correlated with sample size or the percentage of women in a given sample, suggesting that the obtained results are not attributable to sampling bias.

Implications for Gender Differences in Moral Psychology

Our analysis of relative preferences for utilitarian over deontological judgments on incongruent dilemmas replicates earlier findings by Fumagalli and colleagues (2010), suggesting that men have a stronger preference for utilitarian over deontological judgments than women. However, when broken down via PD, it becomes apparent that this difference is almost entirely driven by gender differences in deontological inclinations, with little evidence for gender differences in utilitarian inclinations. To the extent that utilitarian inclinations stem from cognitive evaluations of outcomes and deontological inclinations are shaped by affective responses to harm (Greene et al., 2001), our findings suggest that gender differences in moral judgment are driven by affective responses to harm rather than cognitive evaluations of outcomes. This conclusion is consistent with previous evidence showing that gender differences in affective processing are strong and robust, whereas gender differences in cognitive processing tend to be weak, if they exist at all.

Although the current findings indicate systematic gender differences in responses to moral dilemmas, our conclusions should not be interpreted as involving claims regarding the origin of these differences. Both evolutionary accounts (e.g., Preston & De Waal, 2002) and social learning accounts (e.g., Eagly & Wood, 1999) offer potential reasons why women show stronger deontological inclinations than men. Although the present work provides deeper insights into the exact nature of gender differences in moral judgments, it is agnostic regarding the origin of these differences. That said, the finding that women appear to experience both deontological and utilitarian inclinations may reflect the fact that women in modern societies have to juggle more social roles with potentially conflicting affordances than men (Eagly & Karau, 2002).

Caveats and Limitations

The homogeneity of gender differences across samples corroborates the validity of our conclusions. Nevertheless, there are some limitations that need to be acknowledged. The fact that all 40 studies used the same set of moral dilemmas may raise questions about the generality of our findings. Although PD can help us determine which of two processes is driving a pattern of responding, the results obtained with PD still depend on the materials being used to measure those processes. Thus, the possibility that a different set of moral dilemmas yields different results cannot be ruled out. Future research using PD with different sets of moral dilemmas would help to further corroborate the validity of our conclusions.

In addition, it is important to acknowledge two methodological limitations inherent to nearly all moral dilemma studies. First, moral dilemmas require participants to accept *closed world assumptions* by answering the dilemma as presented, without inserting any new assumptions. For example, the crying baby dilemma requires the assumption that the murderous soldiers will kill the townsfolk and that this outcome is inescapable except by smothering the crying baby. If participants reject such assumptions, then they may reject killing the baby for utilitarian and not deontological reasons. Problems associated with closed world assumptions are not unique to PD but are inherent to all moral dilemma research. Second, incongruent moral dilemmas typically involve an action–inaction confound, such that the choice implied by the utilitarian principle always requires action, whereas the choice implied by the deontological principle always requires inaction. This confound is problematic because harm is often judged worse when it results from action than inaction (Cushman, Young, & Hauser, 2006). Again, this problem is not unique to PD but pertains to most moral dilemma research. We are currently engaged in research designed to overcome this limitation (Gawronski, Conway, Armstrong, Friesdorf, & Hütter, in press).

Finally, every meta-analysis is limited by the breadth and depth of the obtained sample of studies. Typically, there is a concern that meta-analyses may over-estimate the true size of a given effect because studies with nonsignificant results are less likely to be published (Hedges & Vevea, 1996; Rosenthal, 1979). In our meta-analytic reanalysis, however, the file drawer problem seems less of a concern, given that most of the included studies are currently unpublished. Moreover, none of this work was designed to test the effects in question, the only exception being two unpublished studies by Friesdorf and Conway (2013). Thus, rather than an attempt to clarify the strength of a phenomenon across different paradigms, the current work amounts to a secondary analysis of existing datasets. Although the majority of these datasets came from our own lab, it seems unlikely that this restriction affected the obtained results, given that the reported findings did not depend on the type of sample.

Conclusion

Gilligan (1982) famously argued that men prefer a cognitive, abstract, depersonalized method of moral decision making, which she dubbed an *ethic of justice*. Women, in contrast, were claimed to prefer an *ethic of care*, involving moral decisions based on relations to and emotional bonds with particularized others. Although the evidence for Gilligan's claims is rather weak (Jaffee & Hyde, 2000), a lack of gender differences in care and justice orientations should not be equated with a general lack of gender differences in moral psychology. After all, gender differences in non-moral information processing may influence moral judgments in a manner that does not map onto the two kinds of ethics proposed by Gilligan. In line with this conjecture, the current research investigated gender differences in responses to moral dilemmas, showing that men have a stronger preference for utilitarian over deontological judgments than women. PD further indicated that this gender difference is primarily driven by stronger deontological inclinations among women rather than stronger utilitarian inclinations among men. Together with earlier evidence for affective underpinnings of deontological inclinations and cognitive underpinnings of utilitarian inclinations, these findings suggest that women experience stronger affective reactions to harm. However, the current findings cast doubt on the hypothesis that men and women differ in terms of their cognitive evaluations of outcomes. Drawing on the words of James Joyce, both men and women are governed by lines of intellect—women: additionally by curves of emotion.

Appendix

Process Dissociation (PD) Parameter Calculation

To calculate the deontology and utilitarian PD parameters, it is necessary to examine responses to both congruent and incongruent dilemmas. Utilitarianism entails acting in ways that maximize overall outcomes, whereas deontology entails avoiding harmful action regardless of outcomes. Harmful action maximizes overall outcomes in the incongruent, but not in the congruent, dilemmas. These response patterns and their underlying processes are depicted in the processing tree in Figure 1. The top path in the tree illustrates the case where utilitarianism drives the response on a given dilemma; this case entails rejecting harm for congruent dilemmas but accepting harm for incongruent dilemmas. The second path illustrates the case where deontology drives the response on a given dilemma; this case entails rejecting harm for both congruent and incongruent dilemmas. Finally, the bottom path represents the case where neither utilitarianism nor deontology drives the response on a given dilemma; this case entails accepting harm for both congruent and incongruent dilemmas.

Using the two columns on the right side of the figure, it is possible to go backward and determine the cases that lead participants to judge harm as acceptable or unacceptable for congruent and incongruent dilemmas. For congruent dilemmas, harm will be judged as unacceptable either when utilitarianism drives the response, U, or when deontology drives the response, $(1 - U) \times D$. Conversely, harm will be judged as acceptable on congruent dilemmas when neither utilitarianism nor deontology drives the response, $(1 - U) \times (1 - D)$. For incongruent dilemmas, harm will be judged as unacceptable when deontology drives the response, $(1 - U) \times D$. Conversely, harm will be judged acceptable either when utilitarianism drives the response, U, or when neither utilitarianism nor deontology drives the response, $(1 - U) \times (1 - D)$.

The probability of a particular judgment can be algebraically represented as the combination of these cases. For example, the probability of judging harm as unacceptable for congruent dilemmas is represented by the case where either utilitarianism drives responses or deontology drives responses:

$$p(\text{unacceptable} \mid \text{congruent}) = U + \left[(1-U) \times D\right]. \quad \text{(A1)}$$

Conversely, the probability of judging harm as acceptable in congruent dilemmas is represented by the case that neither utilitarianism nor deontology drives responses:

$$p(\text{acceptable} \mid \text{congruent}) = (1-U) \times (1-D). \quad \text{(A2)}$$

For incongruent dilemmas, the probability of judging harm as unacceptable is represented by the case that deontology drives responses:

$$p(\text{unacceptable} \mid \text{congruent}) = \left[(1-U) \times D\right]. \quad \text{(A3)}$$

Conversely, the probability of judging harm as acceptable for incongruent dilemmas is represented by the cases that utilitarianism drives responses, or neither deontology nor utilitarianism drives responses:

$$p(\text{acceptable} \mid \textit{in}\text{congruent}) = U + \left[(1-U) \times (1-D)\right]. \quad \text{(A4)}$$

By algebraically representing the probabilities of accepting and rejecting harm in congruent and incongruent dilemmas, it becomes possible to enter a participants' pattern of actual responses on multiple congruent and incongruent dilemmas, and algebraically combine these equations to solve for two parameters estimating deontological (D) and utilitarian (U) inclinations underpinning their responses. In particular, by including Equation A3 into Equation A1, the latter can be solved for U, leading to the following formula:

$$U = p(\text{unacceptable} \mid \text{congruent}) - p(\text{unacceptable} \mid \text{incongruent}). \quad \text{(A5)}$$

Moreover, by including the calculated value for U in Equation A3, this equation can be solved for D, leading to the following formula:

$$D = \frac{p(\text{unacceptable} \mid \text{incongruent})}{1-U}. \quad \text{(A6)}$$

Together, these formulas enable researchers to obtain parameters that independently estimate the strength of deontological and utilitarian inclinations underlying relative moral dilemma judgments.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by a doctoral fellowship from the Social Sciences and Humanities Research Council of Canada (SSHRC) to the second author (767-2008-1666), and by a grant from the Canada Research Chairs (CRC) Program to the third author (215983).

Note

1. Due to the mathematical structure of the process dissociation (PD) model, estimating the D parameter is not possible when a participant judges harm as acceptable on all incongruent dilemmas and, at the same time, judges harm as unacceptable on all congruent dilemmas. There were a total of 39 participants (0.64%) who accepted causing harm on all 10 incongruent dilemmas; there were 429 (7.03%) who rejected causing harm on all 10 congruent dilemmas. However, not a single participant fell into both of these categories, and it was therefore not necessary to exclude any participants due to this mathematical constraint.

Supplemental Material

The online supplemental material is available at http://pspb.sagepub.com/supplemental.

References

References marked with an asterisk indicate studies included in the meta-analysis.

Amit, E., & Greene, J. D. (2012). You see, the ends don't justify the means: Visual imagery and moral judgment. *Psychological Science*, *23*, 861-868.

*Armstrong, J., Conway, P., & Gawronski, B. (2013). *Moral judgment and socioeconomic status*. Unpublished manuscript.

Bartels, D. M. (2008). Principled moral sentiment and the flexibility of moral judgment and decision making. *Cognition*, *108*, 381-417.

Bartels, D. M., & Pizarro, D. A. (2011). The mismeasure of morals: Antisocial personality traits predict utilitarian responses to moral dilemmas. *Cognition*, *121*, 154-161.

Benhabib, S. (1985). The generalized and the concrete other: The Kohlberg-Gilligan controversy and feminist theory. *PRAXIS International*, *4*, 402-424.

Brody, L. R., & Hall, J. A. (2000). Gender, emotion, and expression. In M. Lewis & J. M. Haviland-Jones (Eds.), *Handbook of emotions: Part IV: Social/personality issues* (2nd ed., pp. 325-414). New York, NY: Guilford Press.

Bullis, C., & Horn, C. (1995). Get a little closer: Further examination of nonverbal comforting strategies. *Communication Reports*, *8*, 10-17.

*Burgmer, P., Forstmann, M., Todd, A.R., & Mussweiler, T. (2015). *Beliefs in mind-body dualism attentuate internal-state reliance*. Manuscript in preparation, University of Cologne, Germany.

Cacioppo, J. T., Petty, R. E., Feinstein, J. A., & Jarvis, W. B. G. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in need cognition. *Psychological Bulletin*, *119*, 197-253.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

*Conway, P., Bartels, D. M., & Pizarro, D. A. (2013). *An exploration of genuinely utilitarian inclinations: Clarifying the relations between moral judgments, antisociality, and prosociality*. Manuscript in preparation, University of Cologne.

*Conway, P., & Gawronski, B. (2013a). *Anger versus sadness: Disambiguating the impact of different emotions on moral judgments*. Unpublished manuscript.

*Conway, P., & Gawronski, B. (2013b). Deontological and utilitarian inclinations in moral decision-making: A process dissociation approach. *Journal of Personality and Social Psychology*, *104*, 216-235.

*Conway, P., & Gawronski, B. (2013c). *Effects of moral and religious priming on moral dilemma judgments*. Unpublished manuscript.

*Conway, P., & Gawronski, B. (2013d). *Effects of vividness, empathy, and emotion regulation on moral judgments*. Unpublished manuscript.

*Conway, P., & Gawronski, B. (2013e). *Moral persuasion: Modifying moral judgments through argumentation*. Unpublished manuscript.

*Conway, P., Love, E., & Mottner, S. (2013). *A four factor model of moral judgments*. Manuscript in preparation, University of Cologne.

*Cross, S. E., & Madson, L. (1997). Models of the self: Self-construals and gender. *Psychological Bulletin*, *122*, 5-37.

Cushman, F., Young, L., & Hauser, M. (2006). The role of conscious reasoning and intuition in moral judgment: Testing three principles of harm. *Psychological Science*, *17*, 1082-1089.

Damian, R. I., & Sherman, J. W. (2013). A process-dissociation examination of the cognitive processes underlying unconscious thought. *Journal of Experimental Social Psychology*, *49*, 228-237.

Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological Review*, *109*, 573-598.

Eagly, A. H., & Wood, W. (1999). The origins of sex differences in human behavior: Evolved dispositions versus social roles. *American Psychologist*, *54*, 408-423.

Eisenberg, N., & Lennon, R. (1983). Sex differences in empathy and related capacities. *Psychological Bulletin*, *94*, 100-131.

Ferreira, M. B., Garcia-Marques, L., Sherman, S. J., & Sherman, J. W. (2006). A dual-process approach to judgment under uncertainty. *Journal of Personality and Social Psychology*, *91*, 797-813.

Fischer, A. H., & Manstead, A. S. R. (2000). The relation between gender and emotion in different cultures. In A. H. Fischer (Ed.), *Gender and emotion: Social psychological perspectives* (pp. 71-96). Cambridge, UK: Cambridge University Press.

Foot, P. (1967). The problem of abortion and the doctrine of double effect. *Oxford Review*, *5*, 5-15.

*Forstmann, M., & Conway, P. (2013). *Arousal and moral judgment*. Unpublished manuscript.

*Friesdorf, R., & Conway, P. (2013). *Gilligan in trolleyland: Resurrecting the debate over gender differences in morality*. Manuscript in preparation, Wilfrid Laurier University.

Fumagalli, M., Ferrucci, R., Mameli, F., Marceglia, S., Mrakic-Sposta, S., Zago, S., . . . Priori, A. (2010). Gender-related differences in moral judgments. *Cognitive Processing*, *11*, 219-226.

*Gamez-Djokic, M., & Molden, D. (2013). *The role of regulatory focus in moral judgment*. Manuscript in preparation, Northwestern University.

Gawronski, B., Conway, P., Armstrong, J., Friesdorf, R., & Hütter, M. (in press). Moral dilemma judgments: Disentangling deontological inclinations, utilitarian inclinations, and general action tendencies. In J. P. Forgas, P. A. M. Van Lange, & L. Jussim (Eds.), *Social psychology of morality*. New York, NY: Psychology Press.

Gilligan, C. (1982). *In a different voice: Psychological theory and women's development*. Cambridge, MA: Harvard University Press.

Gilligan, C., & Attanucci, J. (1988). Two moral orientations: Gender differences and similarities. *Merrill-Palmer Quarterly*, *34*, 223-237.

Greene, J. D. (2007). Why are VMPFC patients more utilitarian? A dual-process theory of moral judgment explains. *Trends in Cognitive Sciences*, *11*, 322-323.

Greene, J. D., Cushman, F. A., Stewart, L. E., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2009). Pushing moral buttons: The interaction between personal force and intention in moral judgment. *Cognition*, *111*, 364-371.

Greene, J. D., & Haidt, J. (2002). How (and where) does moral judgment work? *Trends in Cognitive Science*, *6*, 517-523.

Greene, J. D., Morelli, S. A., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2008). Cognitive load selectively interferes with utilitarian moral judgment. *Cognition*, *107*, 1144-1154.

Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron*, *44*, 389-400.

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, *293*, 2105-2108.

Gross, J. J., & John, O. P. (1998). Mapping the domain of expressivity: Multimethod evidence for a hierarchical model. *Journal of Personality and Social Psychology*, *74*, 170-191.

Hall, J. A., & Schmid Mast, M. (2008). Are women always more interpersonally sensitive than men? Impact of goals and content domain. *Personality and Social Psychology Bulletin*, *34*, 144-155.

Hedges, L. V., & Vevea, J. L. (1996). Estimating effect size under publication bias: Small sample properties and robustness of a random effects selection model. *Journal of Educational and Behavioral Statistics*, *21*, 299-332.

Higgins, J. P. T., Thompson, S. G., Deeks, J. D., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal*, *327*, 557-560.

Hyde, J. S. (1981). How large are cognitive gender differences? A meta-analysis using ω^2 and *d*. *American Psychologist*, *36*, 892-901.

Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist*, *60*, 581-592.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*, 513-541.

Jaffee, S., & Hyde, J. S. (2000). Gender differences in moral orientation: A meta-analysis. *Psychological Bulletin*, *126*, 703-726.

Johnston, D. K. (1988). Adolescents' solutions to dilemmas in fables: Two moral orientations—Two problem solving strategies. In C. Gilligan, J. V. Ward, J. M. Taylor, & B. Bardige (Eds.). *Mapping the moral domain: A contribution to psychological theory and education* (pp. 49-69). Cambridge, MA: Harvard University Press.

Kant, I. (1785/1959). *Foundation of the metaphysics of morals* (L. W. Beck, Trans.). Indianapolis, IN: Bobbs-Merrill.

*Kappes, A., & Van Bavel, J. (2013). *Meta-ethical beliefs shape moral judgments*. Manuscript in preparation, University College London.

Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M., & Damasio, A. (2007). Damage to the prefrontal cortex increases utilitarian moral judgments. *Nature*, *446*, 908-911.

Kohlberg, L. (1969). Stage and sequence: The cognitive-developmental approach to socialization. In D. A. Goslin (Ed.), *Handbook of socialization theory and research* (pp. 347-480). Chicago, IL: Rand McNally.

Krebs, D. L., Vermeulen, S. C., Denton, K. L., & Carpendale, J. I. (1994). Gender and perspective differences in moral judgement and moral orientation. *Journal of Moral Education*, *23*, 17-26.

Lipsey, M. W., & Wilson, D. S. (2001). *Practical meta-analysis* (Vol. 49). Thousand Oaks, CA: SAGE.

Marín-Martínez, F., & Sánchez-Meca, J. (2010). Weighting by inverse variance or by sample size in random-effects meta-analysis. *Educational and Psychological Measurement*, *70*, 56-73.

Meyer, N. D., & Tormala, Z. L. (2010). "Think" versus "feel" framing effects in persuasion. *Personality and Social Psychology Bulletin*, *36*, 443-454.

Mill, J. S. (1861/1998). *Utilitarianism*. In R. Crisp (Ed.), New York, NY: Oxford University Press.

Nichols, S., & Mallon, R. (2006). Moral dilemmas and moral rules. *Cognition*, *100*, 530-542.

Parsons, T., & Bales, R. F. (1955). *Family, socialization and interaction process*. Glencoe, IL: Free Press.

Payne, B. K. (2001). Prejudice and perception: The role of automatic and controlled processes in misperceiving a weapon. *Journal of Personality and Social Psychology*, *81*, 181-192.

Payne, B. K., & Bishara, A. J. (2009). An integrative review of process dissociation and related models in social cognition. *European Review of Social Psychology*, *20*, 272-314.

*Peysakhovich, A., Conway, P., Rand, D., & Greene, J. D. (2013). *Prosocial behavior and moral judgment: Clarifying the relations between economic generosity, deontology, and utilitarianism*. Manuscript in preparation, Harvard University.

Preston, S. D., & De Waal, F. (2002). Empathy: Its ultimate and proximate bases. *Behavioral and Brain Sciences*, *25*, 1-20.

Richard, F. D., Bond, C. F., Jr., & Stokes-Zoota, J. J. (2003). One hundred years of social psychology quantitatively described. *Review of General Psychology*, *7*, 331-363.

*Robinson, J. S., Conway, P., Gawronski, B., & Inzlicht, M. (2013). *Moral judgment and their relation to event-related potentials*. Unpublished manuscript.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*, 638-641.

Skoe, E. E., & Diessner, R. (1994). Ethic of care, justice, identity, and gender: An extension and replication. *Merrill-Palmer Quarterly*, *40*, 272-289.

Strohminger, N., Lewis, R. L., & Meyer, D. E. (2011). Divergent effects of different positive emotions on moral judgment. *Cognition*, *119*, 295-300.

Suter, R. S., & Hertwig, R. (2011). Time and moral judgment. *Cognition*, *119*, 454-458.

*Wisneski, D., Conway, P., & Skitka, L. (2013). *We're not as (politically) different as you think: How do liberals and conservatives moralize the political domain?* Manuscript in preparation, University of Illinois at Chicago.