Prologue to a Multinomial Model of Moral Judgment

Revised title:

Moral Dilemma Judgments: Disentangling Deontological Inclinations, Utilitarian

Inclinations, and General Action Tendencies."

Bertram Gawronski

University of Texas at Austin, USA

Paul Conway

University of Cologne, Germany

Joel Armstrong

University of Western Ontario, Canada

Rebecca Friesdorf

Wilfrid Laurier University, Canada

Mandy Hütter

University of Tübingen, Germany

Introduction

For centuries, societies have wrestled with the question of how to balance the rights of the individual versus the greater good; is it acceptable to ignore a person's rights in order to increase the overall well-being of a larger number of people? The contentious nature of this issue is reflected in many contemporary examples, including debates about whether it is legitimate to cause harm in order to protect societies against threats (e.g., shooting an abducted passenger plane to prevent a terrorist attack) and whether it is acceptable to refuse life-saving support for some people in order to protect the well-being of many others (e.g., refusing the return of American citizens who got infected with Ebola in Africa for treatment in the United States). These issues have captured the attention of social scientists, politicians, philosophers, lawmakers, and citizens alike, partly because they involve a conflict between two moral principles.

The first principle, often associated with the moral philosophy of Immanuel Kant, emphasizes the irrevocable universality of rights and duties. According to the principle of *deontology*, the moral status of an action is derived from its consistency with context-independent norms (*norm-based morality*). From this perspective, violations of moral norms are unacceptable irrespective of the anticipated outcomes (e.g., shooting an abducted passenger plane is always immoral because it violates the moral norm not to kill others). The second principle, often associated with the moral philosophy of John Stuart Mill, emphasizes the greater good. According to the principle of *utilitarianism*, the moral status of an action depends on its outcomes, more specifically its consequences for overall well-being (*outcome-based morality*). From this perspective, violations of moral norms can be acceptable if they increase the well-being of a larger number of people (e.g., shooting an abducted passenger plane is morally

acceptable if it safeguards the well-being of many others). Although both principles are intuitively plausible, their simultaneous consideration can cause feelings of moral conflict when they suggest different conclusions in a particular situation. Over the past decade, research in moral psychology has identified numerous determinants of deontological and utilitarian judgments, thereby providing valuable insights into the psychological processes underlying moral decision-making.

Despite the exponentially growing body of research on deontological and utilitarian judgments, a deeper understanding of their underlying processes has been undermined by two fundamental problems: (1) the treatment of deontological and utilitarian inclinations as opposite ends of a single bipolar continuum rather than independent dimensions, and (2) the conflation of the two moral inclinations with general action tendencies. In the current chapter, we review our ongoing pilot work on a mathematical model that resolves these problems by disentangling and quantifying the unique contributions of (1) deontological inclinations, (2) utilitarian inclinations, and (3) general action tendencies. We argue that this model offers a more fine-grained analysis of the psychological underpinnings of moral judgments, thereby imposing tighter constraints on current theories of moral psychology.

Moral Principles, Moral Judgments, and Psychological Processes

Although research in moral psychology has sometimes conflated normative, empirical, and theoretical aspects of morality (e.g., Kohlberg, 1969), contemporary approaches draw a sharp distinction between (1) moral principles, (2) moral judgments, and (3) underlying psychological processes. Moral principles are abstract philosophical propositions that specify the general characteristics that make an action moral or immoral. According to the principle of deontology, the moral status of an action depends on its consistency with moral norms (e.g., do

not inflict harm upon others). A central aspect of deontology is that the validity of these norms is situation-independent; they always apply regardless of the circumstances. In contrast, the principle of utilitarianism states that the morality of an action depends on its outcomes, in particular its consequences for overall well-being. According to this principle, the context surrounding an action is essential, because the same action may increase well-being in some situations and decrease well-being in others. Thus, unlike the emphasis of context-independent norms in the principle of deontology, the principle of utilitarianism emphasizes the significance of the particular situation. Although the two moral principles often suggest the same conclusion regarding the moral status of an action (e.g., harming a person is immoral because it violates the moral norm not to inflict harm onto others and usually reduces overall well-being), the two principles can lead to conflicting conclusions when an action violates a moral norm, but increases overall well-being (e.g., harming a person is morally acceptable by utilitarian standards, but not by deontological standards, if it protects the lives of many others).

Moral principles have to be distinguished from moral judgments, which may be consistent or inconsistent with a particular principle. For example, to the extent that an empirically observed judgment is consistent with the principle of deontology, it may be described as deontological judgment. Similarly, empirically observed judgments that are consistent with the principle of utilitarianism are often described as utilitarian judgments. A well-known example is Foot's (1967) trolley dilemma, in which a runaway trolley will kill five people unless the trolley is redirected to a different track, causing the death of only one person instead of five. In research using the trolley dilemma, the decision to redirect the trolley is often described as utilitarian, because it maximizes the well-being of a larger number of people. Conversely, the decision not to redirect the trolley is often described as deontological, because it conforms to the

moral norm of not to inflict harm upon others (e.g., Greene, Sommerville, Nystrom, Darley, & Cohen, 2001).

Importantly, the mere consistency of a judgment with a particular moral principle does not imply that the psychological processes underlying the judgment involved the actual use of that principle (Cushman, Young, & Hauser, 2006). In the philosophical literature, this issue is known as the difference between *rule-following* and *rule-conforming* judgments (Wittgenstein, 1953). Whereas rule-following judgments are overt responses that result from the actual application of the relevant rule, rule-conforming judgments are overt responses that are consistent with the rule, but may or may not involve an actual application of this rule in the production of the response. For example, although deontological decisions in the trolley dilemma may stem from the deliberate application of the moral norm not to inflict harm upon others, the mere consistency of the decision with that norm does not imply its actual use in the decision-making process. Over the past decade, the distinction between *rule-following* and *rule-conforming* judgments has become a central theme in moral psychology, in that many theories explain moral judgments in terms of psychological processes that do not involve a reasoned application of moral principles (Greene & Haidt, 2002).

A Dual-Process Theory of Moral Judgment

One of the most prominent examples of such theories is Greene's dual-process theory of moral judgment (Greene et al., 2001). The central assumption of the theory is that deontological and utilitarian judgments have their roots in two distinct psychological processes. Whereas utilitarian judgments are assumed to be the product of controlled cognitive evaluations of outcomes, deontological judgments are assumed to stem from automatic emotional responses to the idea of causing harm. To test these assumptions, moral psychologists have examined

responses to moral dilemmas designed to pit deontology against utilitarianism, such as the trolley dilemma and various structurally similar scenarios (for a review, see Christensen, Flexas, Calabrese, Gut, & Gomila, 2014). Although the unrealistic, comical scenario of the trolley dilemma has raised concerns about its suitability to investigate moral judgments about real-world issues (Bauman, McGraw, Bartels, & Warren, 2014), the evidence obtained with this and structurally similar dilemmas is largely consistent with Greene's dual-process theory.

The hypothesized link between deontological judgments and automatic emotional responses is supported by studies showing increased activation of brain areas associated with emotional processes when participants considered personal moral dilemmas involving direct contact with the victim (Greene et al., 2001) and when participants made deontological judgments on difficult moral dilemmas (Greene, Nystrom, Engell, Darley, & Cohen, 2004). Participants made fewer deontological judgments when emotional distance from victims was increased (Petrinovich, O'Neill, P., & Jorgensen, 1993), after a humorous video clip that presumably reduced negative affect by trivializing the harm dealt to victims (Valdesolo & DeSteno, 2006), or when they suffered damage to emotional brain regions (Ciaramelli, Muccioli, Ladavas, & di Pellegrino, 2007; Koenigs, Young, Adolphs, Tranel, Cushman, Hauser, & Damasio, 2007; Mendez, Anderson, & Shapira, 2005). Conversely, participants made more deontological judgments when imagining harm in vivid detail (Bartels, 2008; Petrinovich & O'Neill, 1996), while experiencing physiological stress (Starcke, Ludwig, & Brand, 2012), and after listening to a morally uplifting story that evoked warm feelings (Strohminger, Lewis, & Meyer, 2011).

The hypothesized link between utilitarian judgments and controlled cognitive processes is supported by studies showing increased activation in brain areas associated with working

memory when participants considered impersonal moral dilemmas in which victims are distant

(Greene et al., 2001) and when participants made utilitarian judgments on difficult dilemmas

(Greene et al., 2004). Facilitating rational decision-making increased utilitarian judgments

(Bartels, 2008; Nichols & Mallon, 2006), whereas introducing time pressure (Suter & Hertwig,

2011) reduced utilitarian judgments, and cognitive load impaired reaction times for utilitarian but

not deontological judgments (Greene, Morelli, Lowenberg, Nystrom, & Cohen, 2008).

Participants with greater working memory capacity were more likely to make utilitarian

judgments (Moore, Clark, & Kane, 2008), as were participants higher in deliberative, as opposed

to intuitive, thinking styles (Bartels, 2008). Together, these findings are consistent with the view

that deontological judgments stem from affective response to the idea of causing harm, whereas

utilitarian judgments stem from cognitive evaluations of outcomes (Greene et al., 2001).

Two Conceptual Problems

Although moral dilemma research has provided many interesting insights into the

determinants of utilitarian versus deontological judgments, the traditional dilemma approach

suffers from two important drawbacks that undermine its suitability for understanding the

psychological underpinnings of moral judgments. The first problem is the treatment of

deontological and utilitarian judgments as opposite ends of a bipolar continuum, which stands in

contrast to the assumption that they are rooted in functionally independent processes (see

Conway & Gawronski, 2013). In the traditional dilemma approach, participants must categorize

harmful action as either acceptable or unacceptable, thereby making a judgment that conforms to

either the deontological or the utilitarian principle. To behave in line with the deontological

principle is to simultaneously behave in opposition to the utilitarian principle, and vice versa.

Thus, the traditional approach confounds *selecting* one option with *rejecting* the other. This

confound would be acceptable if the moral inclinations underlying overt judgments were themselves inversely related (i.e., stronger inclinations of one kind are associated with weaker inclinations of the other kind). However, a central assumption of Greene's dual-process theory is that deontological and utilitarian judgments stem from two functionally independent processes, thereby allowing for the possibility that the two moral inclinations are active at the same time. Indeed, the entire field of moral dilemma research is predicated on the assumption that people experience a psychological conflict when the two moral inclinations suggest different courses of action. Such conflict would not occur if the two inclinations were inversely related. The significance of this problem is illustrated by the fact that any empirical finding (e.g., difference in moral judgments across conditions) can be attributed to either (1) differences in deontological inclinations, (2) differences in utilitarian inclinations, or (3) differences in both. An illustrative example is a study by Bartels and Pizarro (2011) showing that psychopaths tend to make more utilitarian judgments compared to non-psychopathic participants. However, counter to interpretations of this effect as reflecting stronger utilitarian inclinations among psychopaths, it seems more plausible that psychopaths have no concerns about violating moral norms, rather than strong concerns with maximizing the well-being of others. Such ambiguities in the interpretation of empirical findings undermine not only the possibility of drawing strong theoretical conclusions regarding the psychological underpinnings of moral judgments; they also diminish the value of these findings for practical applications outside of the lab.

A second major problem of the traditional dilemma approach is that it conflates the two moral inclinations with general preferences for action versus inaction (van den Bos, Müller, & Damen, 2011). In the classic dilemma approach, the utilitarian choice always requires action, whereas the deontological choice always requires inaction. However, preferences for action

versus inaction may differ for various reasons that are unrelated to deontological and utilitarian inclinations (Abarracin, Hepler, & Tannenbaum, 2011; Carver & Scheier, 1994; Kuhl, 1985). Distinguishing between genuine deontological inclinations and general action aversion is important, because deontological concerns can sometimes suggest action rather than inaction (e.g., bringing an American citizen who got infected with Ebola in Africa to the United States for treatment). Although this possibility has been largely ignored in moral dilemma research, it plays a central role in research on proscriptive versus prescriptive morality (e.g., Janoff-Bulman, Sheikh, & Hepp, 2009). Whereas *proscriptive norms* specify what people should not do, *prescriptive norms* specify what people should do. Although harm caused by action is often perceived as more immoral than equivalent harm caused by inaction (e.g., Cushman et al., 2006; Spranca, Minsk, & Baron, 1991), the principle of deontology—defined as norm-based morality—implies that both actions and inactions can be immoral if they conflict with a moral norm. Whereas actions are immoral if they conflict with a proscriptive norm (e.g., pushing someone in front of a car), inactions are immoral if they conflict with a prescriptive norm (e.g., not helping the victim of a car accident).

Similar concerns apply to the confound between utilitarianism and action, because a general preference for action can produce a "utilitarian" judgment in situations where the utilitarian principle suggests action (i.e., moral dilemmas involving proscriptive norms) and a "deontological" judgment in situations where the deontological principle suggests action (i.e., moral dilemmas involving prescriptive norms). Such action tendencies have to be distinguished from utilitarian inclinations, which involve a genuine concern to maximize well-being. An illustrative example is the finding that people with high levels of testosterone show a greater willingness to inflict harm upon one person to increase the well-being of several others (Carney

& Mason, 2010), which may be due to stronger utilitarian inclinations, weaker deontological

inclinations, or both. Yet, an alternative interpretation is that individuals with high levels of

testosterone simply have a stronger tendency to act regardless of whether action is consistent

with the principle of deontology or utilitarianism (see Andrew & Rogers, 1972; Joly, Alibhai,

Galicia, Park, Yi, Wagner, & Tannock, 2006; Lynn, Houtman, Weather, & Ketterson, 2000).

Thus, similar to the non-independence of deontological and utilitarian inclinations in the

traditional dilemma approach, the confound between the two moral inclinations and general

action tendencies poses a major challenge for unambiguous interpretations of empirical results,

thereby undermining the possibility of drawing strong theoretical and practical conclusions.

Process Dissociation as a Solution to the Non-independence Problem

To overcome the first problem—the non-independence of deontological and utilitarian

judgments—Conway and Gawronski (2013) developed a process dissociation (PD) model to

disentangle and quantify the independent contributions of deontological and utilitarian

inclinations to overt moral judgments. Although originally designed to examine memory

(Jacoby, 1991), 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). The key to PD is employing 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 dilemma research, incongruent dilemmas pit deontological

against utilitarian inclinations, such that harmful action is acceptable from a utilitarian view, but

unacceptable from a deontological view (e.g., causing harm increases overall well-being).

Congruent dilemmas have structure and wording identical to incongruent dilemmas, except that

dealing harm leads to worse outcomes overall (e.g., causing harm does not increase overall well-

being). Thus, harmful action is unacceptable by either deontological or utilitarian moral

standards. For example, in the incongruent version of the vaccine dilemma (see first column of

Table 1), a doctor must decide whether to administer a vaccine with potential deadly side-effects

in order to cure an even deadlier disease, thereby saving many lives. In the congruent version of

the vaccine dilemma (see second column of Table 1), a doctor must decide whether to administer

a vaccine with potential deadly side-effects to cure the common flu, thereby reducing discomfort

but not saving lives. Thus, harmful action maximizes overall welfare in the incongruent, but not

in the congruent version, of the vaccine dilemma.

To illustrate the logic underlying PD, participants' judgments in congruent and

incongruent moral dilemmas can be illustrated by means of a processing tree (see Figure 1). Each

path from left to right depicts judgment outcomes on the two kinds of dilemmas as a function of

distinct underlying processes. The three paths in the figure capture the three cases that (1)

utilitarianism drives the response on a given dilemma (top path), (2) deontology drives the

response (middle path), and (3) neither utilitarianism nor deontology drives the response (bottom

path). U depicts the case that utilitarianism drives the response, and D depicts the case that

deontology drives the response. Conversely, $1 - U$ depicts the case that utilitarianism does not

drive the response, and $1 - D$ depicts the case that deontology does not drive the response.

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 example, on congruent dilemmas, harmful action will

be judged as unacceptable when utilitarianism drives the response (U). Alternatively, if

utilitarianism does not drive the response ($1 - U$), harmful action will be judged as unacceptable

when deontology drives the response (D). Harmful action will be judged as acceptable in

congruent dilemmas only when neither utilitarianism ($1 - U$) nor deontology ($1 - D$) drives the response. Similarly, on incongruent dilemmas, participants will judge harmful action as unacceptable when utilitarianism does not drive the response ($1 - U$) and, at the same time, deontology does drive the response (D). However, harmful action will be judged as acceptable either when utilitarianism drives the response (U), or alternatively when neither utilitarianism ($1 - U$) nor deontology ($1 - D$) drives the response.

By means of the processing paths depicted in Figure 1, it is now possible to create mathematical equations that delineate the probability of a particular overt judgment on congruent and incongruent dilemmas as a function of the two underlying inclinations. For example, the probability of judging harmful action as unacceptable on congruent dilemmas is represented by the cases where (1) utilitarianism drives the response, and (2) deontology drives the response when utilitarianism fails to drive the response. In algebraic terms, this probability is represented by the equation:

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

Conversely, the probability of judging harmful action as acceptable on congruent dilemmas is represented by the case that neither utilitarianism nor deontology drives the response, which can be represented algebraically as:

$$(2)\ p(\text{acceptable} \mid \text{congruent}) = (1 - U) \times (1 - D).$$

The same logic can be applied to incongruent dilemmas. For example, the probability of judging harmful action as unacceptable on incongruent dilemmas is represented by the case that deontology drives the response when utilitarianism does not drive the response. Algebraically, this likelihood is represented by the equation:

$$(3)\ p(\text{unacceptable} \mid \text{incongruent}) = (1 - U) \times D.$$

Conversely, the probability of judging harmful action as acceptable on incongruent dilemmas is represented by the cases that (1) utilitarianism drives the response, and (2) neither deontology nor utilitarianism drives the response. In algebraic terms, this probability is represented as:

$$(4)\ p(\text{acceptable} \mid \text{incongruent}) = U + [(1 - U) \times (1 - D)].$$

Using the empirically observed probabilities of participants' *acceptable* and *unacceptable* responses on congruent and incongruent dilemmas, these equations can be used to calculate numerical estimates for the two kinds of moral tendencies by solving them algebraically for the two parameters representing deontology (D) and utilitarianism (U).[1] Specifically, by including Equation 3 into Equation 1, the latter can be solved for U, leading to the following formula:

$$(5)\ U = p(\text{unacceptable} \mid \text{congruent}) - p(\text{unacceptable} \mid \text{incongruent}).$$

Moreover, by including the calculated value for U in Equation 3, this equation can be solved for D, leading to the following formula:

$$(6)\ D = p(\text{unacceptable} \mid \text{incongruent}) / (1 - U).$$

These two formulas allow researchers to quantify the strength of deontological and utilitarian inclinations within participants by using their individual probabilities of showing a particular response on the two kinds of moral dilemmas. The resulting parameter values can then be used as measurement scores in experimental designs to investigate differences across conditions and in correlational designs to investigate relations to individual difference or

---

[1] Note that Equation 1 and 2 are mathematically redundant, because $p(\text{acceptable} \mid \text{congruent}) = 1 - p(\text{unacceptable} \mid \text{congruent})$. Similarly, Equation 3 and 4 are mathematically redundant, because $p(\text{acceptable} \mid \text{incongruent}) = 1 - p(\text{unacceptable} \mid \text{incongruent})$. Thus, the basic logic of PD is to solve two (non-redundant) equations for two unknowns on the basis of observed data.

criterion measures (for a more detailed discussion of technical details of PD, see Appendix B of Conway & Gawronski, 2013).

In their original application of PD to moral dilemma responses, Conway and Gawronski (2013) found that individual differences in perspective taking and empathic concern were positively related to D, but not U. Conversely, individual differences in need for cognition were positively related to U, but not D. Moreover, individual differences in moral identity were positively related to both D and U, a pattern that was concealed in the traditional approach due to the treatment of the two moral inclinations as opposite ends of a bipolar continuum. Two experimental studies further showed that cognitive load reduced U without affecting D, whereas increased salience of harm increased D without affecting U. Together, these results demonstrate not only the usefulness of PD to disentangle and quantify the functionally independent contributions of deontological and utilitarian inclinations to moral dilemma judgments; they also provide more compelling support for Greene et al.'s (2001) dual-process theory, suggesting that deontological judgments are rooted in automatic emotional responses to the idea of causing harm, whereas utilitarian judgments are the product of controlled cognitive evaluations of outcomes.

A Multinomial Model of Moral Judgment

Although Conway and Gawronski's (2013) PD model provides a solution to the first problem—the non-independence of deontological and utilitarian judgments—it does not resolve the second problem, because it retains the confound between the two moral inclinations and general action tendencies. D scores still conflate deontological inclinations with a general preference for inaction, and U scores still conflate utilitarian inclinations with a general preference for action. To simultaneously resolve both conceptual problems of traditional

dilemma research, we recently developed an extended model that provides separate parameters for (1) deontological inclinations, (2) utilitarian inclinations, and (3) action aversion (see Figure 2). To emphasize the conceptual and stochastic difference to the parameters of Conway and Gawronski's PD model, the three parameters are depicted with the two-digit acronyms *De* (for deontology), *Ut* (for utilitarianism), and *In* (for inaction). The central difference to Conway and Gawronski's PD model is that the extended model captures cases in which the deontological principle prohibits action (i.e., proscriptive dilemmas) as well as cases in which the deontological principle prescribes action (i.e., prescriptive dilemmas). For either type of dilemma, the moral implication of the utilitarian principle depends on the respective outcomes, such that action is acceptable in proscriptive dilemmas and inaction is acceptable on prescriptive dilemmas if either decision increases overall well-being. Thus, the parameter estimates of the extended model are based on participants' responses to four kinds of moral dilemmas that differ with regard to whether (1) the dilemma involves a proscriptive or prescriptive norm and (2) the outcomes of action versus inaction suggest utilitarian choices that are either congruent or incongruent with the deontological norm (for an example, see Table 1). Because the three processes lead to different outcomes on the four kinds of dilemmas (see Figure 2), the extended model allows us to disentangle and quantify their unique contributions to moral dilemma judgments, thereby resolving the two conceptual problems of the traditional approach.

Although the derivation of the model equations follows the same logic described for Conway and Gawronski's (2013) PD model, there are a few important differences in the mathematical underpinnings of the two models. Different from the use of linear algebra in the calculation of the two PD scores, our extended model uses multinomial modeling to estimate parameter values for the three processes (see Batchelder & Riefer, 1999). Whereas PD is based

on two (non-redundant) equations with two unknowns, multinomial modeling involves a higher number of equations than unknowns. Thus, whereas PD scores can be calculated directly by means of linear algebra, parameter estimations in multinomial modeling are based on maximum likelihood statistics. Specifically, multinomial modeling involves systematic adjustments in the parameter values to minimize the differences between the actual probabilities of observed responses and the probabilities predicted by the model. The deviation between actual and predicted probabilities serves as the basis for statistical tests of goodness-of-fit, which provides evidence regarding the validity of the model in describing the data. If the deviation between actual and predicted probabilities is small, fit statistics will reveal a non-significant deviation between the two, suggesting that the model accurately describes the data. If, however, the deviation between actual and predicted probabilities is large, fit statistics will reveal a significant deviation between the two, indicating that the model does not accurately describe the data. To the extent that the model fits the data, the parameter estimates can be used to investigate effects of experimental manipulations and correlations with individual difference or criterion measures, similar to the PD approach (for an example, see Conrey, Sherman, Gawronski, Hugenberg, & Groom, 2005).

Preliminary Findings

To test the validity of our multinomial model, we conducted a pilot study in which participants were asked to indicate for a set of newly created moral dilemmas whether the decision suggested in the dilemma is acceptable or unacceptable. The dilemmas included 4 parallel versions of 10 different scenarios that varied in terms of whether (1) the dilemma involved a proscriptive or prescriptive norm and (2) the outcomes of action versus inaction suggested utilitarian choices that were either congruent or incongruent with the deontological

norm (for an example, see Table 1). Half of the participants completed congruent and

incongruent versions of the dilemmas involving proscriptive norms; the remaining half

completed congruent and incongruent versions of the dilemmas involving prescriptive norms.

The final sample of our pilot study included 292 participants on Amazon's Mturk.[2] The model fit

the data well, $G^2(1) = 1.28$, $p = .28$. Both the *De* and the *Ut* parameter differed significantly from

zero, demonstrating that both processes contributed participants' responses to our moral

dilemmas (see Table 2). The *In* parameter differed significantly from its reference point of 0.5,

which reflects an equal distribution of *action* and *inaction* tendencies. The finding that the *In*

parameter was significantly higher than zero demonstrates that, on average, participants showed

a general reluctance to act regardless of the dilemma context (see Table 2). Together, these

results indicate that all three processes significantly contributed to overt judgments in the moral

dilemmas, providing preliminary evidence for the validity of our multinomial model.

To further explore the usefulness of our model in providing deeper insights into the

psychological underpinnings of moral judgments, we also investigated gender differences in the

three parameters. A recent meta-analysis ($N = 6,100$) using Conway and Gawronski's (2013) PD

model suggests that women show stronger deontological inclinations than men ($d = .57$), while

men show only slightly stronger utilitarian inclinations than women ($d = .10$) (Friesdorf,

Conway, & Gawronski, 2014). Using our multinomial model, we replicated this pattern in our

pilot study (123 women, 169 men). Whereas women showed significantly higher *De* scores than

men, there were no significant gender differences on the *Ut* parameter (see Figure 3). Yet, our

---

[2] The original sample included 343 participants. Forty-one participants started the study but did not complete it. Ten participants failed to pass an instructional attention check, and were therefore excluded from the analysis (Oppenheimer, Meyvis, & Davidenke, 2009).

extended model also revealed a significant difference on the *In* parameter, in that women showed a significantly stronger aversion to action than men. This result suggests that gender differences in moral dilemma judgments are mostly due to differences in deontological inclinations and action aversion, but not utilitarian inclinations.

Expanding on the results of our pilot study, two follow-up studies aimed to provide deeper insights into the psychological processes underlying deontological inclinations, utilitarian inclinations, and action aversion. A central assumption of Greene et al.'s (2001) dual-process theory is that deontological judgments stem from automatic emotional processes, whereas utilitarian judgments are the product of controlled cognitive processes. Although these assumptions are consistent with a considerable body of research, the available evidence remains ambiguous due to (1) the non-independent measurement of the two moral inclinations in the traditional dilemma approach and (2) the conflation of the two moral inclinations with general action tendencies. For example, it is possible that automatic emotional processes contribute to the rejection of harmful action, not by increasing deontological concerns with norm violations, but by increasing action aversion (Miller, Hannikainen, & Cushman, 2014). Similarly, one could argue that controlled cognitive processes contribute not only to utilitarian assessments of outcomes, but also to deontological assessments of norm violations.

To provide deeper insights into the psychological underpinnings of deontological inclinations, utilitarian inclinations, and action aversion, we asked 288 participants on Amazon's Mturk to indicate for our new set of moral dilemmas whether the described decision is

acceptable or unacceptable.[3] To investigate the resource-dependence of the underlying psychological processes, half of the participants were asked to rehearse 8-digit letter strings while reading and responding to the dilemmas. The remaining half were not asked to rehearse any letter strings. As with our pilot study, our extended model fit the data very well, $G^2(2) = 0.31$, $p = .85$. Consistent with Greene's assumption that utilitarian inclinations are the product of controlled cognitive process, cognitive load significantly reduced the *Ut* parameter (see Figure 4). However, counter to Greene's theory, cognitive load also reduced the *De* parameter, suggesting that deontological inclinations are the product of a cognitively effortful process.

Because the obtained effect of cognitive load on deontological inclinations challenges one of the most central assumptions in moral dilemma research, we aimed to replicate it in a regular lab study with 100 psychology undergraduates. Again, our extended model fit the data very well, $G^2(2) = 1.25$, $p = .53$. Corroborating the validity of the obtained results, cognitive load reduced both the *Ut* parameter and the *De* parameter, indicating that both utilitarian assessments of outcomes and deontological assessments of norm violations involve cognitively effortful processes (see Figure 5).

In our ongoing research, we are exploring the possibility that automatic emotional processes influence overt moral judgments by influencing general action tendencies rather than deontological assessments of norm violations (see Miller et al., 2014). Together with the identified contribution of cognitively effortful processes to deontological assessments of norm

---

[3] The original sample included 335 participants. Thirty participants started the study but did not complete it. Seventeen participants failed to pass an instructional attention check, and were therefore excluded from the analysis (Oppenheimer et al., 2009).

violations, evidence for this hypothesis would require significant revisions in the interpretation of previous findings, posing a major challenge to existing theories of moral judgment.

Conclusion

The current chapter reviewed our ongoing pilot work on a multinomial model of moral judgment. Although previous research provided interesting insights into the determinants of deontological and utilitarian judgments, a deeper understanding of their underlying processes has been undermined by (1) the treatment of deontological and utilitarian inclinations as opposite ends of a single bipolar continuum rather than independent dimensions, and (2) the conflation of the two moral inclinations with general action tendencies. Our multinomial model resolves both conceptual problems by quantifying the unique contributions of (1) deontological inclinations, (2) utilitarian inclinations, and (3) general action tendencies. A major aspect of this endeavor is the integration of both proscriptive and prescriptive norms, the latter of which have been largely ignored in traditional moral dilemma research. By offering a more fine-grained analysis of the psychological underpinnings of moral judgment, our model not only imposes tighter constraints on current theories of moral psychology; it also offers valuable practical insights for the resolution of moral controversies in society.

References

Albarracin, D., Hepler, J., & Tannenbaum, M. (2011). General action and inaction goals: Their behavioral, cognitive, and affective origins and influences. *Current Directions in Psychological Science, 20,* 119-123.

Andrew, R. J., & Rogers, L. J. (1972). Testosterone, search behavior and persistence. *Nature, 237,* 343-345.

Bartels, D. (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.

Batchelder, W. H., & Riefer, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin & Review*, *6,* 57-86.

Bauman, C. W., McGraw, A. P., Bartels, D. M., & Warren, C. (2014). Revisiting external validity: Concerns about trolley problems and other sacrificial dilemmas in moral psychology. *Social and Personality Psychology Compass, 8,* 536-554.

Carney, D. R., & Mason, M. F. (2010). Moral decisions and testosterone: When the ends justify the means. *Journal of Experimental Social Psychology, 46,* 668-671.

Carver, C. S., & Scheier, M. F. (1998). *On the self-regulation of behavior*. New York: Cambridge University Press.

Christensen, J. F., Flexas, A., Calabrese, M., Gut, N. K., & Gomila, A. (2014). Moral judgment reloaded: A moral dilemma validation study. *Frontiers in Psychology, 5:607*.

Ciaramelli, E., Muccioli, M., Ladavas, E., & di Pellegrino, G. (2007). Selective deficit in

personal moral judgment following damage to ventromedial prefrontal cortex. *Social

Cognitive and Affective Neuroscience, 2*, 84–92.

Conrey, F. R., Sherman, J. W., Gawronski, B., Hugenberg, K., & Groom, C. (2005). Separating

multiple processes in implicit social cognition: The quad-model of implicit task

performance. *Journal of Personality and Social Psychology, 89*, 469-487.

Conway, P., & Gawronski, B. (2013). Deontological and utilitarian inclinations in moral

decision-making: A process dissociation approach. *Journal of Personality and Social

Psychology, 104,* 216-235.

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.

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

15.

Friesdorf, R., Conway, P., & Gawronski, B. (2014). *Gender differences in responses to moral

dilemmas: A process dissociation analysis*. Manuscript submitted for publication.

Greene, J. D., & Haidt, J. (2002). How (and where) does moral judgment work? *Trends in

Cognitive Sciences, 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.

Janoff-Bulman, R., Sheikh, S., & Hepp, S. (2009). Proscriptive versus prescriptive morality:

    Two faces of moral regulation. *Journal of Personality and Social Psychology, 96,* 521-

    537.

Joly, F., Alibhai, S. M. H., Galica, J., Park, A., Yi, Q. L., Wagner, L., & Tannock, I. F. (2006).

    Impact of androgen deprivation therapy on physical and cognitive function, as well as

    quality of life of patients with nonmetastatic prostate cancer. *The Journal of Urology*,

    176, 2443-2447.

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.

Kuhl, J. (1985). *Action control: From cognition to behavior*. Berlin: Springer.

Lynn, S. E., Houtman, A. M., Weathers, W. W., Ketterson, E. D., & Nolan Jr, V. (2000).

    Testosterone increases activity but not daily energy expenditure in captive male dark-

    eyed juncos, Junco hyemalis. *Animal Behaviour*, *60,* 581-587.

Mendez, M. F., Anderson, E., & Shapira, J. S. (2005). An investigation of moral judgment in

    frontotemporal dementia. *Cognitive and Behavioral Neurology, 18*, 193-197.

Miller, R. M., Hannikainen, I. A., & Cushman, F. A. (2014). Bad actions or bad outcomes?

Differentiating affective contributions to the moral condemnation of harm. *Emotion, 14,*

573-587.

Moore, A. B., Clark, B. A., & Kane, M. J. (2008). Who shalt not kill? Individual differences in

working memory capacity, executive control, and moral judgment. *Psychological*

Science, 19, 549-57.

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

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks:

Detecting satisficing to increase statistical power. *Journal of Experimental Social*

Psychology, 45, 867-872.

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.

Petrinovich, L., & O'Neill, P. (1996). Influence of wording and framing effects on moral

intuitions. *Ethology & Sociobiology, 17*, 145-171.

Petrinovich, L., O'Neill, P., & Jorgensen, M. (1993). An empirical study of moral intuitions:

Toward an evolutionary ethics. *Journal of Personality and Social Psychology, 64*, 467-

478.

Spranca, M., Minsk, E., & Baron, J. (1991). Omission and commission in judgment and choice.

Journal of Experimental Social Psychology, 27, 76–105.

Starcke, K., Ludwig, A., & Brand, M. (2012). Anticipatory stress interferes with utilitarian moral

judgment. *Judgment and Decision Making, 7*, 61-68.

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.

Valdesolo, P. & DeSteno, D. (2006). Manipulations of emotional context shape moral judgment.

    Psychological Science, 17, 476–477.

van den Bos, K., Müller, P. A., & Damen, T. (2011). A behavioral disinhibition hypothesis of

    interventions in moral dilemmas. *Emotion Review, 3,* 281-283.

Wittgenstein, L. (1953). *Philosophische Untersuchungen*. Oxford: Blackwell.

Proscriptive Dilemma		Prescriptive Dilemma	
Incongruent	Congruent	Incongruent	Congruent
You are the head doctor in a health clinic overrun by patients with a deadly virus. You just received a shipment of drugs that can cure the disease but the drugs have their own severe side-effects. If you administer the drugs to your patients, a small number will die from the side-effects but most will live. If you do not, most will die from the virus. Is it appropriate for you to administer the drug in order to cure the disease, even though some will die from the side-effects?	You are the head doctor in a health clinic overrun by patients with the latest flu virus. You just received a shipment of drugs that can cure the disease but the drugs have their own severe side-effects. If you administer the drugs to your patients, a small number will die from the side-effects but most will live. If you do not, most will continue to suffer from the effects of the virus. Is it appropriate for you to administer the drug in order to cure the disease, even though some will die from the side-effects?	You are the head doctor in a health clinic overrun by patients with a deadly virus. You just received a shipment of drugs that can cure the disease but the drugs have their own severe side-effects. Your patients are slated to receive the drugs later today unless you intervene. If they receive the drugs, a small number of patients will die from the side-effects but most will live. If not, most will die from the virus. Is it appropriate for you to intervene in order to prevent the deadly side-effects, even though most of your patients will die from the disease?	You are the head doctor in a health clinic overrun by patients with the latest flu virus. You just received a shipment of drugs that can cure the disease but the drugs have their own severe side-effects. Your patients are slated to receive the drugs later today unless you intervene. If they receive the drugs, a small number of patients will die from the side-effects but most will live. If not, most will continue to suffer from the effects of the virus. Is it appropriate for you to intervene in order to prevent the deadly side-effects, even though most of your patients will suffer from the disease?

Table 1. Example of a moral dilemma involving either a proscriptive or a prescriptive norm, suggesting deontological decisions that are either congruent or incongruent with utilitarian assessments of outcomes.

Parameter	Estimated Score	Standard Error	95% Confidence Interval
Ut	0.344	0.012	0.320 – 0.368
De	0.228	0.018	0.192 – 0.264
In	0.527	0.012	0.504 – 0.551

Table 2. Parameter estimates for utilitarian inclinations (*Ut*), deontological inclinations (*De*), and action aversion (*In*).
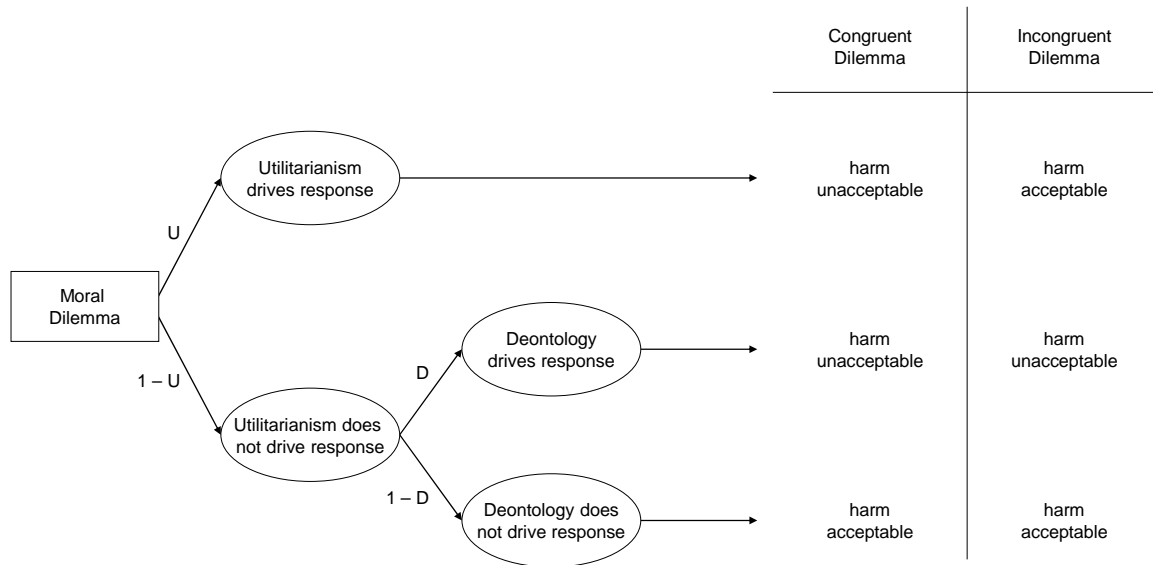
Figure 1. Processing tree illustrating the underlying components leading to judgments that harmful action is either acceptable or unacceptable in congruent and incongruent moral dilemmas. Copyright © 2013 by the American Psychological Association. Reproduced with permission. The official citation that should be used in referencing this material is Conway, P., & Gawronski, B. (2013). Deontological and utilitarian inclinations in moral decision-making: A process dissociation approach. *Journal of Personality and Social Psychology, 104,* 216-235. The use of APA information does not imply endorsement by APA.
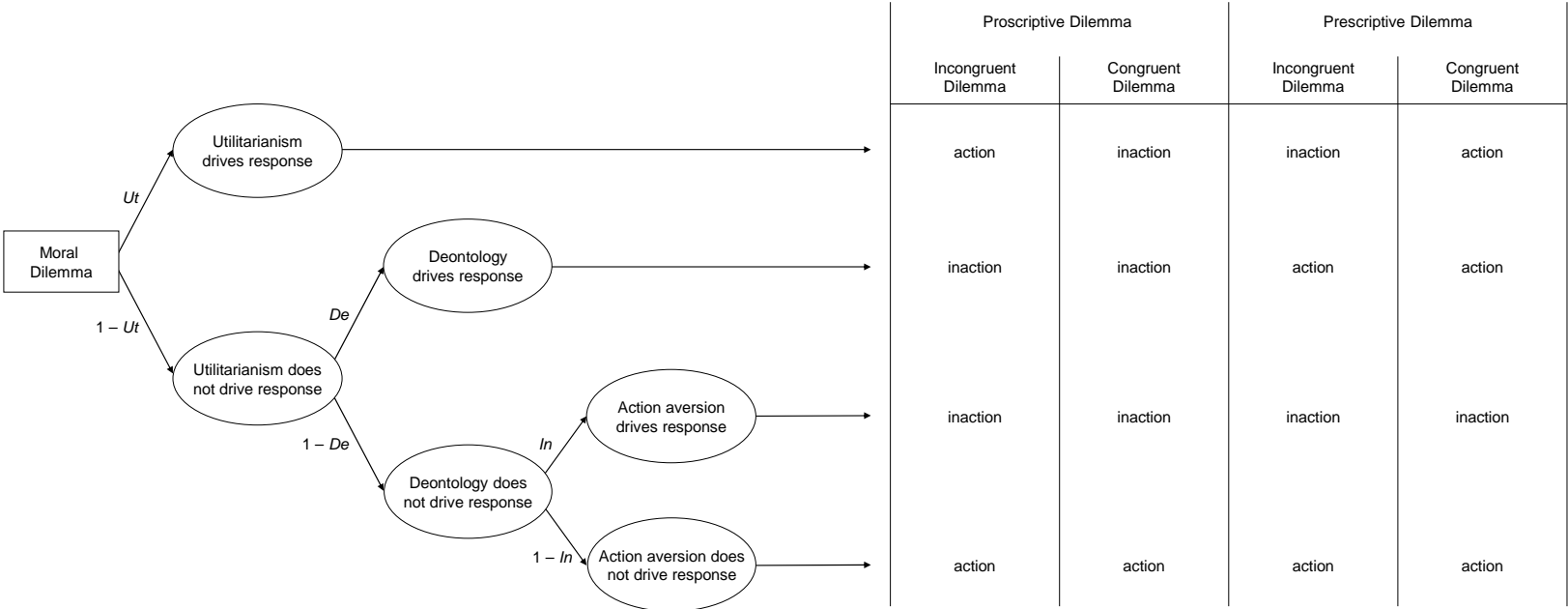
	Proscriptive Dilemma		Prescriptive Dilemma	
	Incongruent Dilemma	Congruent Dilemma	Incongruent Dilemma	Congruent Dilemma
Utilitarianism drives response	action	inaction	inaction	action
Deontology drives response	inaction	inaction	action	action
Action aversion drives response	inaction	inaction	inaction	inaction
Action aversion does not drive response	action	action	action	action

Figure 2. Processing tree illustrating the underlying components leading to action or inaction in congruent and incongruent moral dilemmas involving either proscriptive or prescriptive norms.
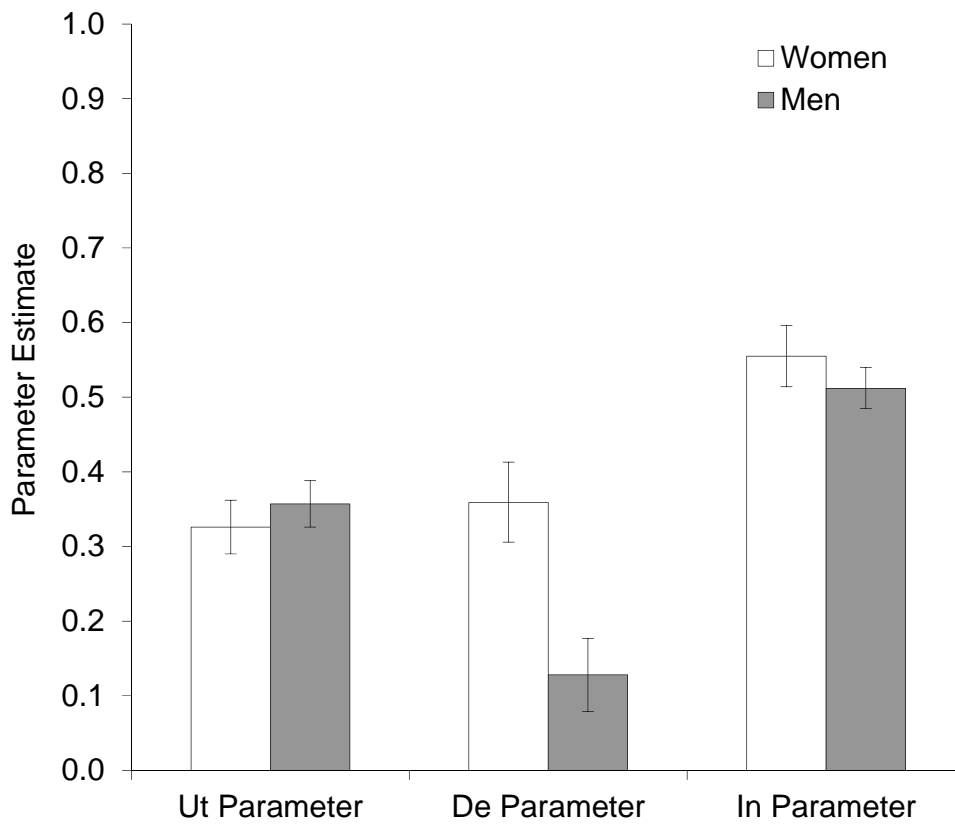
Figure 3. Parameter estimates for utilitarian inclinations (*Ut*), deontological inclinations (*De*),

and action aversion (*In*) for women and men (online study with $N = 292$). Error bars depict 95%
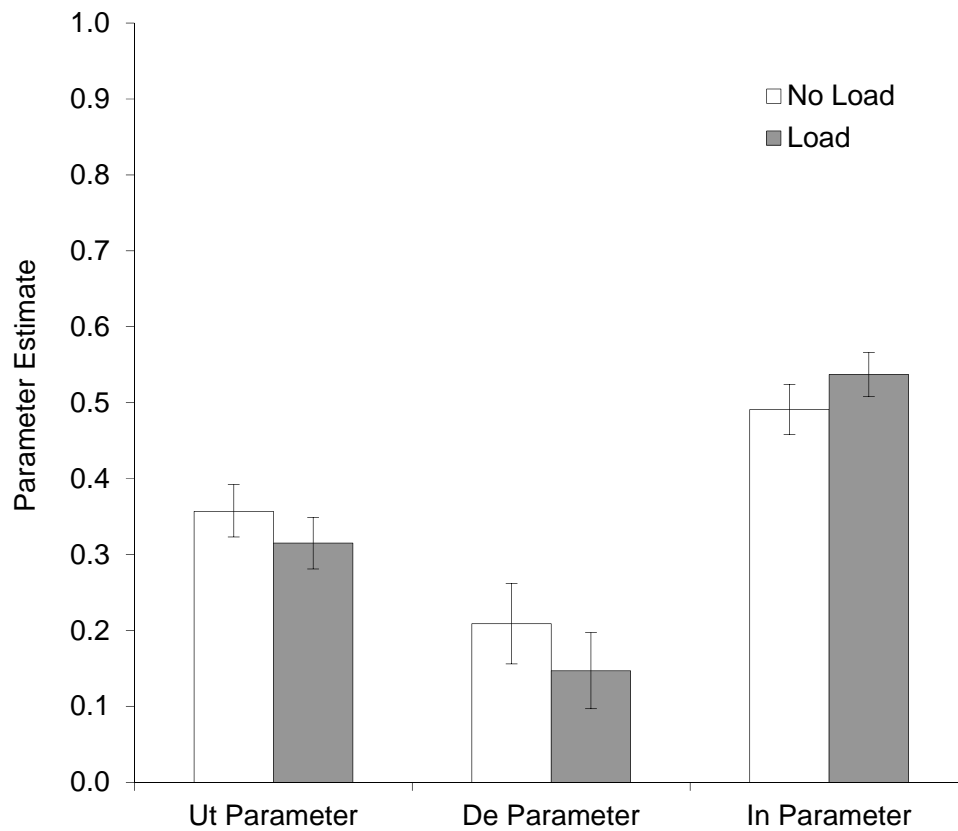
confidence intervals.

Figure 4. Parameter estimates for utilitarian inclinations (*Ut*), deontological inclinations (*De*), and action aversion (*In*) as a function of cognitive load (online study with $N = 288$). Error bars depict 95% confidence intervals.
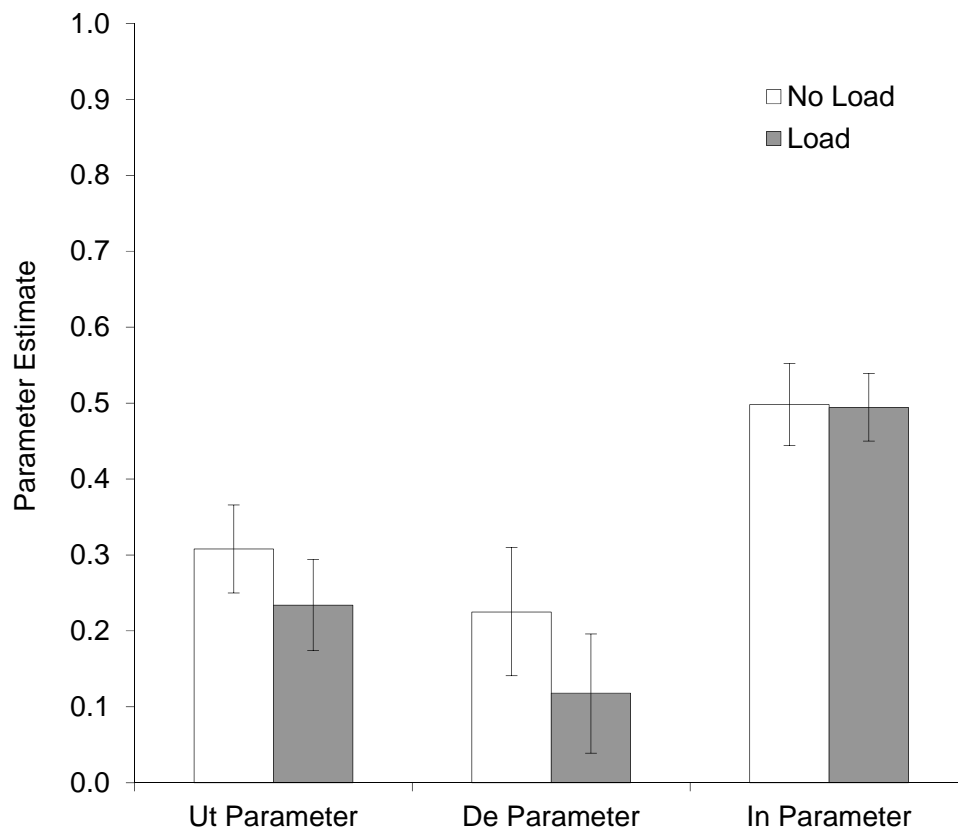
Figure 5. Parameter estimates for utilitarian inclinations (*Ut*), deontological inclinations (*De*), and action aversion (*In*) as a function of cognitive load (lab study with *N* = 100). Error bars depict 95% confidence intervals.