Who Shalt Not Kill?

Individual Differences in Working Memory Capacity, Executive Control, and Moral Judgment

Adam B. Moore
Princeton University

Brian A. Clark and Michael J. Kane
University of North Carolina at Greensboro

Running Head: Working memory and moral judgment

Please address all correspondence to:

Adam B. Moore
Department of Psychology
Green Hall
Princeton University
Princeton, NJ 08550
Voice: 609.258.5108
E-mail: amtwo@princeton.edu

WORD COUNT:
Main text (3843) + Footnotes (90) + Acknowledgments (19) = Total (3952); (30 refs)
Abstract

Findings from Greene and colleagues (2001, 2004) suggest that exerting executive control influences responses to moral dilemmas. Here, subjects judged the moral appropriateness of killing one person to save others in 24 dilemmas, each with dual resolutions balancing physical directness of killing, personal risk to the subject, inevitability of the death, and intentionality of the action. All dilemma variables demonstrate main effects. Executive control was indexed by scores on working memory capacity (WMC) tasks. People with greater WMC found killing more appropriate than those with lesser WMC, under certain circumstances, and were more consistent in their judgments. We also report novel interactions among dilemma variables, such as between more deliberative and more intuitively processed factors, thus challenging existing dual-process frameworks of moral judgment.
When Pilate saw that he could prevail nothing, but that rather a tumult was made, he took water, and washed his hands before the multitude, saying, I am innocent of the blood of this just person… (Matthew 27:24)

The synoptic gospels are notoriously ambiguous in assigning earthly culpability for Jesus’ death. Pilate, rather than actively sentencing Jesus to crucifixion, passively allows it by letting the multitude release someone. When they choose Barabbas, Pilate claims his hands are clean regarding Jesus’ fate. We suggest that the gospel authors captured something important here about humankind’s moral sense, particularly regarding the directness of killing.

Philosophers have long considered such issues in developing prescriptive ethics, but scientists are now studying moral judgments in order to understand what such judgments actually are, and why. A lively theoretical debate concerns the role of emotion versus normative-rule-based cognition in driving moral judgments (e.g., Blair, 1995; Haidt, 2001; Lakoff, 2002; Mikhail, 2007; Nichols, 2002; Pizarro & Bloom, 2003; Prinz, 2006). Much of the excitement surrounding this debate derives from Greene and colleagues’ neuroimaging work (Greene, Nystrom, Engell, Darley, & Cohen, 2004; Greene, Sommerville, Nystom, Darley, & Cohen, 2001; see also Borg, Hynes, Van Horn, Grafton, & Sinnott-Armstrong, 2006). Using fMRI they investigated the neural substrates of judging resolutions to hypothetical personal versus impersonal dilemmas. In the former, subjects contemplate (a) causing serious bodily harm or death to (b) another person or persons, (c) in a way that does not deflect preexisting harm; personal dilemmas thus involve “up-close-and-personal” harmful acts authored by the subject. Impersonal dilemmas do not meet these criteria. A prototypical personal dilemma is the “Footbridge” problem, where a runaway trolley hurtles toward five unaware workmen; the only way to save them is to push a heavy man (standing nearby on a footbridge) onto the track, where he will die stopping the trolley. In an impersonal version, the “Trolley” problem, you may save
the workmen by throwing a switch that sends the trolley onto another track, killing one worker. Most people find the impersonal resolution to be morally more acceptable than the personal, despite their identical consequences (e.g., Hauser, Cushman, Young, Jin, & Mikhail, 2007).

Why? Emotion prevails. Greene and colleagues (2001, 2004) demonstrated that personal dilemmas engage brain regions involved in emotion (e.g., amygdala, posterior cingulate/precuneus), whereas impersonal dilemmas activate areas involved in deliberative reasoning and working memory (e.g., middle frontal gyrus, bilateral parietal lobe). Moreover, when subjects judge personal resolutions to be morally appropriate, they respond slowly and engage networks associated with executive control (e.g., anterior/dorsolateral prefrontal cortex, anterior cingulate). These authors thus propose a dual-process model in which moral personal dilemmas evoke an automatic, “hot” emotional response that biases responding against harm unless “cold” cognitive control overrides it.

Here, then, we tested dual-process theory by asking whether individual differences in control, indexed by working memory capacity (WMC), would predict “cold” consequentialist responses to personal dilemmas. Research with non-emotive tasks indicates that: 1) WMC variation predicts executive ability to override prepotent responses (e.g., Kane, Bleckley, Conway, & Engle, 2001; Kane & Engle, 2003); 2) WMC variation predicts abstract reasoning performance (e.g., Kane, Hambrick, & Conway, 2005; Oberauer, Schulze, Wilhelm, & Süß, 2005) and; 3) experimentally loading WMC impairs reasoning about future consequences (e.g., Hinson, Jameson, & Whitney, 2003). We thus hypothesized that people with greater WMC should more rationally evaluate consequences within personal moral dilemmas by controlling emotion and engaging deliberative processing. This is a risky prediction because executive theories of WMC (e.g., Engle & Kane, 2004; Hasher, Lustig, & Zacks, 2007) are supported
almost exclusively by studies of “cold” cognition (but see Braver, Gray, & Burgess, 2007), and individual differences in such moral judgments are rare (Hauser et al., 2007; O’Neill & Petrinovich, 1998).

Our second goal was to test whether the personal-impersonal distinction would survive a re-tooling of dilemma stimuli. To do this, we redesigned Greene et al.’s (2001, 2004) materials to address significant problems: 1) more personal than impersonal dilemmas involved death or injury; more impersonal than personal dilemmas involved lying or stealing; 2) many impersonal dilemmas presented abstract, probabilistic-reasoning problems; 3) several scenarios posed non-dilemmas (e.g., should a child murder his grandmother for not buying him a gift?); 4) personal dilemmas were longer than impersonal (Ms = 124 and 102 words, respectively); 5) unsystematically, only some dilemmas endangered subjects’ lives; 6) subjects saw multiple versions of some scenarios, allowing carry-over effects (potentially problematic for similar studies: Hauser et al., 2007; Mikhail, 2007; Petrinovich et al., 1993; Valdesolo & DeSteno, 2006).

Some of these shortcomings reflect vagueness in the Greene et al. (2001, 2004) definition of impersonal dilemmas. Because personal dilemmas met 3 criteria (bodily harm; to persons; not deflecting existing harm), any dilemmas meeting 0 – 2 criteria were considered impersonal, allowing wildly different dilemmas to be contrasted with personal ones and thus failing to capture the vexing psychological distinction between Footbridge and Trolley problems. For example, dilemmas involving whether to make charitable donations, steal or damage property, or enact environmental-hazard policies were included with impersonal dilemmas involving whether to deflect impending death threats (e.g., a runaway trolley; poisonous fumes) from an innocent group toward an innocent individual. In order to better equate personal and impersonal dilemmas
for severity of outcomes, we redefined the personal-impersonal distinction to reflect more versus less direct killing (see also Royzman & Baron, 2002; Spranca, Minsk, & Baron, 1991). Here, all critical dilemmas involved killing someone to save more people, but personal dilemmas required more physically direct killing that was less mediated through mechanical/technological means, or through other people’s actions, than did impersonal dilemmas.

Our third goal was to test the influence of other philosophically relevant variables on moral judgment (perhaps moderating personal-impersonal effects). First, because everyday moral dilemmas often involve costs and benefits to the self, and because trolley-dilemma judgments change with subjects’ relations to the hypothetical victims (Petrinovich et al., 1993), our materials systematically varied self-risk. We predicted that killing to save oneself and others would be more acceptable than killing to save only others; we view this self-other distinction as evolutionarily ingrained and thus automatically and intuitively processed (cf. Petrinovich et al.). Second, we manipulated whether the intended object of harm would die regardless of subjects’ actions, and we predicted that subjects would find killing more acceptable if it merely hastened an inevitable death; we view this inevitable-avoidable distinction as relatively subtle and thus more rationally and deliberatively processed. Third, along with others (Borg et al., 2006; Cushman, Young, & Hauser, 2006; Hauser et al., 2007; Mikhail, 2007), we examined Aquinas’ doctrine of double effect (DDE), whereby harm is more permissible as a foreseen but unintended consequence than as an intended means to an end. Normative sensitivity to DDE may partially explain the personal-impersonal distinction: in trolley-type dilemmas, one either intentionally uses the bystander’s body or unintentionally sacrifices another workman. People may therefore blanch at personal-dilemma resolutions because they internalized the DDE. Thus, half our
impersonal dilemmas required killing as a means, and half as an unintended but foreseeable outcome.

METHODS

Subjects

Strict exclusion criteria anticipated future neuroimaging research: 129 native English-speaking undergraduates, who completed WMC screening and reported no history of psychiatric diagnoses and no current illness or (psychiatric or flu) medication, completed the study as a course requirement.

WMC Screening

Subjects individually completed 3 automated complex-span measures: operation span (OSPA), reading span (RSPAN), and symmetry span (SSPAN), which presented short lists of memory items. Each item succeeded a processing task with a response deadline tailored to each task and subject, based upon latencies ($M + 2.5 \text{ SDs}$) during 15 processing-only items (Unsworth, Heitz, Schrock, & Engle, 2005). OSPAN’s processing task was verifying equations involving two operations. In RSPAN, subjects verified sentences’ meaningfulness. In SSPAN, subjects judged matrix patterns for symmetry. In all tasks, each processing stimulus appeared until response or deadline, followed 200 ms later by the memory item (for 250 ms in OSPAN and RSPAN; 650 ms in SSPAN); all were followed by the next processing stimulus or memory test.

OSPA and RSPAN presented lists of 3 – 7 letters to remember from a pool of 12; SSPAN presented 2 – 5 red-square locations to remember within a $4 \times 4$ matrix. In OSPAN and RSPAN, the memory test for each list presented all 12 letters in the same locations and subjects mouse-clicked on the presented letters in serial order. The SSPAN test presented a $4 \times 4$ matrix
and subjects clicked the previously occupied squares in serial order. All tasks presented 3 trials at each list length in random order.

Span scores (total items recalled in serial position; Conway et al., 2005) correlated with $r$s = .70, .54, and .64 ($\text{OSPAN} \times \text{RSPAN}$, $\text{OSPAN} \times \text{SSPAN}$, and $\text{RSPAN} \times \text{SSPAN}$, respectively). We created a WMC score for each subject by averaging $z$-scores for each task; it was normally distributed (skew = -0.497; kurtosis = -0.452).

Moral Judgment Task

Materials

We created dilemmas based on the contextual and consequential identity exemplified by the Greene et al. (2001, 2004) *Footbridge* (Personal) and *Trolley* (Impersonal) scenarios, modifying their materials whenever possible into Critical and Filler dilemmas (for materials, see http://www.uncg.edu/~mjkane/memlab.html; for examples, see Table 1). The subject was the protagonist in all dilemmas, which consisted of an introductory paragraph and a 4- or 5-sentence resolution. All subjects saw the same 14 Fillers, which were similar to Critical dilemmas, but 12 involved no deaths, 1 called for killing many to save 1, and 1 involved killing for personal gain. Fillers thus likened the stimulus list to Greene et al. (2001, 2004).

Our 24 critical scenarios had 2 resolutions each: personal and impersonal. Subjects saw only one version of each dilemma (counterbalanced across subjects). All critical dilemmas involved killing at least 1 person in order to save more, pitting the imperative to maximize good consequences against that to avoid causing harm. Wherever possible, we used identical phrasing for personal and impersonal resolutions, equating them for deaths and matching length to within 2 words.
Subjects’ lives were at risk in 12 of the 24 dilemmas, so killing saved oneself and others (Self); subjects were not in jeopardy in the other 12 (Other). Additionally, in 12 of the 24 dilemmas, the sacrificed lives would be lost regardless of subjects’ actions (Inevitable); in the other 12, they would be lost only if subjects acted (Avoidable). Our materials reflected a $2 \times 2 \times 2$ nested design, with 3 personal and 3 impersonal resolutions nested within 6 Self-Inevitable, Self-Avoidable, Other-Inevitable, and Other-Avoidable dilemmas. Mean sentence and word counts were matched (e.g., $M$ words = 100.9, 95.3, 93.3, 101.2, respectively, for Self-Inevitable, Self-Avoidable, Other-Inevitable, Other-Avoidable). We created 2 lists, each presenting 1 resolution (Personal or Impersonal) per dilemma, with the other list presenting the complement. To test the influence of the DDE on judgments, 6 Impersonal resolutions required killing others as a means to save more (Instrumental) and 6 as a foreseen but unintended consequence (Incidental). All Personal dilemmas were Instrumental.

Procedure

We tested all subjects individually. A computer monitor presented dilemmas in black against grey. The experimenter read aloud on-screen instructions encouraging subjects not to consider legality, but rather moral appropriateness. Subjects completed 2 practice dilemmas before beginning experimental trials. We randomized dilemma order for each subject.

Each introductory paragraph appeared alone until subjects pressed a key to view the resolution sentences, each of which advanced via key-press (all text remained visible until final response). The final sentence asked whether the resolution was morally appropriate; subjects responded via bimanual key-press. The screen blanked for 1 s between dilemmas. After the experiment, subjects completed a demographic/health questionnaire.
RESULTS

All null hypothesis significance tests were non-directional with $\alpha = .05$. $\eta^2_p$ indicates effect size; $p_{rep}$ indicates the probability of replicating effects’ directions given similar procedures and subjects (Killeen, 2005).

Subjects

We excluded data from 16 subjects: 12 ignored instructions, 3 had invalid WMC data, and 1 was agitated. We report data from 113 subjects (68 females).

Moral Judgments

Dilemma Responses

Figure 1 presents mean proportions of affirmative responses (i.e., that killing was morally appropriate) by experimental condition, where Personal-Impersonal, Self-Other, and Inevitable-Avoidable variables all affected judgments as predicted. A repeated-measures ANOVA confirmed that impersonal killing was more appropriate than personal, $F(1,112) = 48.05$, $\eta^2_p = .30$, $p_{rep} \geq .99$, killing to save oneself and others was more appropriate than to save only others, $F(1,112) = 16.88$, $\eta^2_p = .13$, $p_{rep} \geq .99$, and killing someone whose death was inevitable was more appropriate than someone whose death was avoidable, $F(1,112) = 31.35$, $\eta^2_p = .22$, $p_{rep} \geq .99$. Only the 3-way interaction among variables was highly replicable, $F(1,112) = 8.29$, $\eta^2_p = .07$, $p_{rep} = 0.98$.

We first explored this interaction with a 2 (Personal-Impersonal) x 2 (Inevitable-Avoidable) ANOVA on Self dilemmas. The interaction, $F(1,112) = 4.13$, $\eta^2_p = .04$, $p_{rep} = .96$, demonstrated that the Personal-Impersonal distinction was stronger for inevitable than avoidable deaths: When one’s life was at risk and victims’ deaths were inevitable, impersonal killing was more appropriate than when deaths were avoidable, $t(112) = 5.04$, $\eta^2_p = .185$, $p_{rep} \geq .99$. Personal
killing did not elicit this effect, $t(112) = 1.23$, $p_{rep} = .81$. The parallel ANOVA on Other dilemmas yielded the opposite pattern, $F(1,112) = 5.24$, $\eta_p^2 = .05$, $p_{rep} = .94$. Here, when saving only others, impersonal killing was equally appropriate regardless of inevitability of victims’ deaths, $t(112) = 1.61$, $p_{rep} = .87$, but personal killing was more acceptable when deaths were inevitable, $t(112) = 4.72$, $\eta_p^2 = .17$, $p_{rep} \geq .99$. Note also that the two Personal-Self conditions and Personal-Other-Inevitable did not differ (all $t$’s $< 1.23$, all $p_{rep}$’s $< .81$), indicating that, whereas either self-threat or inevitability of death, alone, attenuated aversion to personal killing, they had no combined effect. In contrast, when impersonal killing was required, Self and Inevitable variables interacted to yield particularly high “appropriate” judgments under self-risk and inevitable death (i.e., Impersonal-Self-Inevitable $>$ Impersonal-Self-Avoidable and Impersonal-Other-Inevitable, $t$s $> 3.65$, $p_{reps} \geq .99$; the latter were equivalent, $t(112) = .419$, $p_{rep} = .62$).

To address the DDE’s influence, we tested Impersonal dilemmas for effects of instrumental versus incidental killing. Impersonal killing was more appropriate when it involved the unintentional-but-foreseen death of another rather than killing as means to an end (Ms = .635 and .585, respectively), $t(112) = 2.33$, $\eta_p^2 = .05$, $p_{rep} = .95$. However, impersonal-instrumental killing was more appropriate than personal killing overall (Ms = .585 and .472, respectively), $t(112) = 4.79$, $p_{rep} \geq .99$; thus, the DDE cannot account for the Personal-Impersonal distinction.

**WMC**

We assessed the role of WMC in moral judgment by re-conducting our analyses with WMC as a covariate (treating WMC as a continuous variable; Conway et al., 2006; Oberauer, 2005). The only effect involving WMC was its interaction with Personal-Impersonal and Inevitable-Avoidable, $F(1,111) = 3.43$, $\eta_p^2 = .03$, $p_{rep} = .90$. Figure 2 presents the means for low,
medium, and high WMC subjects by tertiary split (collapsed over Self-Other); WMC-groups’ judgments look different only for personally hastening inevitable deaths.

We pursued this interaction with an ANCOVA for Personal dilemmas, yielding a WMC × Inevitable-Avoidable interaction, $F(1,111) = 3.28, \eta_p^2 = .03, p_{rep} = .90$. When considering killing someone who would die anyway, higher WMC subjects found personal killing more appropriate than did lower WMC subjects; all subjects found personal killing equally inappropriate for avoidable deaths. The parallel ANCOVA on Impersonal dilemmas found no WMC effects, $F$s < 1. Thus, only higher WMC subjects accounted for inevitability of death in judging appropriateness of personal killing. (An ANCOVA on Impersonal-Instrumental versus Impersonal-Incidental dilemmas to test the DDE found no WMC effects, $F$s < 1.)

A final WMC analysis tested whether better cognitive control predicted greater consistency of judgment, especially for Personal dilemmas, which more strongly challenged cognitive control. For each subject we scored each “appropriate” and “inappropriate” response as 1 and 0, respectively, and calculated SD across judgments. Because WMC interacted with Personal-Impersonal and Inevitable-Avoidable conditions above, we collapsed SDs over Self-Other. ANCOVA indicated that higher WMC subjects were less variable than were lower WMC subjects, $F(1,111) = 4.29, \eta_p^2 = .04, p_{rep} = .93$, and WMC affected variability differently in Personal and Impersonal dilemmas, $F(1,111) = 3.48, \eta_p^2 = .03, p_{rep} = .90$ (for other WMC effects, $F$s < 1). Follow-up analyses verified that, as illustrated in Figure 3, WMC affected consistency across Personal, $F(1,111) = 7.55, \eta_p^2 = .064, p_{rep} = .97$, but not Impersonal dilemmas ($F < 1$).
Response Time (RT)

A 2 (Personal-Impersonal) × 2 (Self-Other) × 2 (Inevitable-Avoidable) repeated measures ANOVA\(^3\) on mean RTs indicated that Personal dilemmas were judged faster than Impersonal dilemmas (Ms = 4783 and 5601 ms, respectively), \(F(1,110) = 18.62, \eta^2_p = .15, p_{rep} \geq .99\), but Self-Other and Inevitable-Avoidable had no main effects (\(F_s < 1\)). Personal-Impersonal interacted with Inevitable-Avoidable, \(F(1,110) = 4.18, \eta^2_p = .04, p_{rep} = .92\): For personal killing, inevitable deaths yielded slower judgments than did avoidable ones (Ms = 4999 and 4566 ms, respectively), \(t(110) = 2.29, \eta^2_p = .05, p_{rep} = .94\), suggesting some deliberation was elicited to reconcile a bias against personal killing with that toward hastening inevitable death. For impersonal killing, inevitable and avoidable scenarios yielded equivalent RTs (Ms = 5415 and 5788 ms, respectively), \(t(110) = -1.22, p_{rep} = .81\). Regarding possible DDE effects, RTs did not differ between Impersonal-Instrumental and Impersonal-Incidental killing (Ms = 5844 and 6221 ms, respectively; 5 subjects excluded for empty cells), \(F(1,107) = 1.44, p_{rep} = .80\).

We also attempted to replicate Greene et al.’s (2001) Stroop-like finding of slowed “appropriate” responses to Personal versus Impersonal dilemmas (and versus “inappropriate” responses to Personal dilemmas). To do so, we treated Response as a factor in a 2 (Personal-Impersonal) × 2 (Appropriate-Inappropriate) ANOVA (10 subjects excluded for empty cells). “Appropriate” and “inappropriate” RTs were equivalent (see Table 2; \(F < 1\)) and Response did not interact with Personal-Impersonal, \(F(1,102) = 1.07, p_{rep} = .77\). We thus failed to replicate Greene et al.’s (2001) central behavioral result.

To examine our new variables for RT effects involving Response, we first conducted a 2 (Self-Other) × 2 (Appropriate-Inappropriate) ANOVA that yielded only a significant interaction, \(F(1,97) = 6.97, \eta^2_p = .07, p_{rep} \geq .99\) (15 subjects excluded for empty cells). When subjects were
at risk, “appropriate” responses were significantly faster than “inappropriate”, $t(98) = -2.20$, $p_{rep} = .94$ (see Figure 4; Table 2). However, when only others’ lives were at risk, subjects were slower to endorse than to denounce killing, $t(108) = 1.97$, $p_{rep} = .92$. Thus, subjects responded most quickly to save themselves and to abandon others; in contrast, sacrificing themselves and saving others required longer consideration. Regarding the inevitability factor, a 2 (Inevitable-Avoidable) × 2 (Appropriate-Inappropriate) ANOVA yielded only a significant interaction, $F(1,101) = 6.95$, $\eta_p^2 = .064$, $p_{rep} > .99$ (11 subjects excluded for empty cells). In Avoidable dilemmas, subjects were slower to endorse than to denounce killing (see Table 2), $t(106) = 1.73$, $p_{rep} = .89$, suggesting some deliberation over whether to take a life that would otherwise be spared; there was a weak effect, in the opposite direction, for Inevitable dilemmas (see Table 2), $t(104) = 1.34$, $p_{rep} = .83$.

Finally, Greene et al.’s (2004) dual-process model predicts that higher WMC subjects should be faster than lower WMC subjects to endorse Personal killing, as higher WMC subjects better control Stroop-like response conflicts (e.g., Kane & Engle, 2003). A 2 (Personal/Impersonal) × 2 (Appropriate/Inappropriate) ANCOVA with WMC as the covariate found no support for the predicted 3-way interaction (or for a WMC main effect; both $F$s < 1; 6 subjects excluded for empty cells). Indeed, in direct contradiction to dual-process models, greater WMC predicted slightly longer RTs for “appropriate” responses to Personal dilemmas, $r = .18$, $p_{rep} = .90$ (all other $r$s < .10, $p_{rep}s < .75$).

**DISCUSSION**

Hypothetical moral judgments about killing someone in order to save others are affected by the direct, personal nature of inflicted harm, by the benefit to the agent and the inevitability of victims’ deaths, and by individual differences in WMC. Our subjects found “personal,” direct
harm to be especially inappropriate when it saved only others rather than themselves and others, and when the harmed parties would not otherwise be killed. In contrast, they found “impersonal,” indirect harm to be particularly appropriate to save themselves and when the harmed parties would be killed regardless. Moreover, individual differences in WMC, which reflect cognitive-control variation, predicted moral judgments when personal killing harmed those whose fate was already sealed, as well as the consistency with which people made moral judgments about personal harm.

Uniquely, these findings reinforce the importance and domain generality of WMC beyond typical laboratory tasks, demonstrate individual differences in moral judgment based on executive control capability (despite noteworthy failures to identify individual differences in these judgments; Hauser et al., 2007), and clarify the particular roles of executive control in moral judgment. Although we replicated Greene et al.’s (2001, 2004) Personal-Impersonal distinction while narrowing its definition to relative directness, our data challenge the idea of executive control overriding prepotent emotional processes as necessary to produce ‘appropriate’ responses to provocative dilemmas. First, WMC did not interact with the Personal-Impersonal variable alone, as we would expect if cognitive control simply restrains emotional reactions. Higher WMC subjects were most likely to endorse personal killing only when harm was inevitable, which seems unlikely to have resulted simply from executive control over emotion, but rather from deliberative reasoning. The association between higher WMC and longer RTs when responding “appropriate” to personal killing supports this interpretation, as does the greater consistency in higher WMC subjects’ judgments of personal killing (consistency across similar cases is a hallmark of rationality/reasoning; e.g. von Neumann & Morgenstern, 1947). Second, we did not replicate Greene et al.’s (2001) Stroop-like findings of slow affirmative responses to
Personal dilemmas. Instead, our subjects, with our improved stimulus materials, responded faster to Personal than Impersonal dilemmas, regardless of judgment.

Our WMC results may reflect a selectively engaged, voluntary reasoning system that follows only after successful executive control (DeNeys, 2006). Although this suggestion requires that WMC did not predict judgments elsewhere (e.g., Self-Personal, Other-Personal) because self threat was processed automatically, we created the Self-Other variable with this interpretation in mind. Although further research must determine whether Self-Other is processed intuitively, our data provide preliminary support, with more and faster “appropriate” responses and fewer and slower “inappropriate” responses to Self dilemmas, and the reverse for Other dilemmas. The lack of this pattern and the interaction with WMC suggest that inevitability is processed more effortfully/deliberately. The present findings are important because, if Self-Other is processed automatically, they suggest direct cooperation between some types of automatic/ emotional processing and deliberation/reasoning in generating moral judgments, outside of (or subsequent to) executive override functions: Self-Other and Inevitable-Avoidable variables had interactive effects only for Impersonal dilemmas, not for Personal. A simple egoistic bias, alone, cannot explain why self-preservation and inevitability combined to yield higher appropriateness judgments only when subjects considered impersonal harm. Thus, automatically and more deliberatively processed variables may influence the formation of moral judgments in their convergence, and not only in their conflict (cf. Greene et al., 2004).
References


Footnotes

1. Many moral-judgment studies rely exclusively on subtle variations of the trolley problem (e.g., Hauser, Cushman, Young, Jin, & Mikhail, 2007; Petrinovich, O’Neill, & Jorgensen, 1993; Valdesolo & DeSteno, 2006).

2. One dilemma (“Modified Fumes”) was miscoded and deleted prior to analyses.

3. Data from 3 outliers were deleted. Response (“appropriate” versus “inappropriate”) was not included in this omnibus ANOVA because most subjects had no observations in at least one cell. Subsequent ANOVAs test for Response effects in Personal-Impersonal, Self-Other, and Inevitable-Avoidable conditions separately, where fewer subjects had missing cells.
Acknowledgment

This work was supported by Brian Clark’s UNCG Research Assistantship. We acknowledge Shannon Holmes’s data-collection assistance and helpful comments.
Figure Captions

Figure 1. Mean proportion of “appropriate” judgments to moral-dilemma resolutions, by dilemma type (Personal/Impersonal harm, Self/Other benefic, and Inevitable/Avoidable death). Error bars depict standard errors.

Figure 2. Mean proportion of “appropriate” judgments by working memory capacity (WMC) group (a tertiary split on the WMC composite variable) and dilemma type (Personal/Impersonal harm and Inevitable/Avoidable death) collapsed over the Self/Other variable. Error bars depict standard errors. Note: High = High WMC group; Med = Medium WMC group; Low = Low WMC group.

Figure 3. Mean standard deviation across moral judgments by WMC group and Personal/Impersonal dilemma type. Error bars depict standard errors. Note: High = High WMC group; Med = Medium WMC group; Low = Low WMC group.

Figure 4. Mean response time (in milliseconds) to dilemmas by Self/Other and response, collapsed over Personal/Impersonal and Inevitable/Avoidable. Note: Appropriate = “Yes/Appropriate” responses; Inappropriate = “No/Inappropriate” responses.
Figure 1

This figure illustrates the proportion of "appropriate" responses for different scenarios. The x-axis represents the categories of scenarios: Self-Inevitable, Self-Avoidable, Other-Inevitable, and Other-Avoidable. The y-axis shows the proportion of "appropriate" responses, ranging from 0 to 0.8.

Bars for Personal and Impersonal responses are depicted, with error bars indicating variability. The data suggests a higher proportion of "appropriate" responses for Personal scenarios compared to Impersonal ones across all categories.
Figure 2

![Graph showing the proportion of "Appropriate" responses for different conditions: Personal-Inevitable, Impersonal-Inevitable, Personal-Avoidable, and Impersonal-Avoidable. The graph includes bars for Low, Med, and High conditions.](image-url)
Figure 3

The bar graph illustrates the mean SD of judgments for personal and impersonal scenarios across different working memory conditions: Low, Med, and High. The y-axis represents the Mean SD of Judgments, while the x-axis categorizes the scenarios as Personal and Impersonal. The graph shows the variability in judgments under each condition.
Figure 4.
Table 1. Sample dilemmas.

1) Self-Benefic, Avoidable Death

You and five other people are trapped in a burning building. There is only one emergency exit through which all of you could escape to safety, but it is blocked by burning debris. You notice another person in the hallway leading to the exit who has been injured but is about to crawl to safety through a small hole at the bottom of the exit door. You and the five people behind you do not have time to climb through the small hole.

a) Personal resolution

You realize that you could grab the injured man and use his body as a battering-ram to break through the burning blockage in the hallway that is preventing your escape. Doing this is certain to kill him. However, if you do not do this, you and the five people behind you will die. Is it appropriate for you to kill this man in order to save yourself and the other five people?

b) Impersonal resolution

The hallway’s emergency system puts out fire by eliminating oxygen from the hall and you can activate the system by pressing a nearby button. The fire will go out, but the injured person will suffocate and die. However, if you do not do this, you and the five people behind you will die. Is it appropriate for you to activate the system in order to save yourself and the other five people?

2) Other-benefic, Inevitable Death

You are a worker in a mine. The only way to exit the mine is to ride up in rock-buckets that can hold up to three people at a time. It is the end of the workday and miners from lower levels are riding up past you. As you are waiting for a bucket to ride in, you notice that the cable
supporting the rock buckets is about to snap. If the cable snaps, all of the miners in the buckets will fall to their deaths.

\textit{a) Personal resolution}

The only way to prevent this is to use your axe to hit the last bucket causing it to flip over and dump its contents, lightening the load enough to save the miners above. There is one miner in this bucket who will be killed as a result. Is it appropriate to flip over the last miner’s bucket, killing him, to save the other miners?

\textit{b) Impersonal resolution}

The only way to prevent this is to hit the emergency bucket release switch which will automatically detach the last bucket from the cable, lightening the load just enough to save the miners above. There is one miner in this bucket who will be killed as a result. Is it appropriate to detach the last miner’s bucket, killing him, to save the other miners?
Table 2. Mean reaction times by factor level and response.

<table>
<thead>
<tr>
<th>Factor Level</th>
<th>Response</th>
<th>‘Appropriate’</th>
<th>‘Inappropriate’</th>
</tr>
</thead>
<tbody>
<tr>
<td>Personal</td>
<td>‘Appropriate’</td>
<td>4987 ms (2611)</td>
<td>4817 ms (2503)</td>
</tr>
<tr>
<td></td>
<td>‘Inappropriate’</td>
<td>5520 ms (2912)</td>
<td>5794 ms (3117)</td>
</tr>
<tr>
<td>Impersonal</td>
<td>‘Appropriate’</td>
<td>4988 ms (2417)</td>
<td>5706 ms (3508)</td>
</tr>
<tr>
<td></td>
<td>‘Inappropriate’</td>
<td>5512 ms (2697)</td>
<td>4983 ms (2291)</td>
</tr>
<tr>
<td>Self</td>
<td>‘Appropriate’</td>
<td>5021 ms (2225)</td>
<td>5696 ms (3191)</td>
</tr>
<tr>
<td></td>
<td>‘Inappropriate’</td>
<td>5374 ms (3102)</td>
<td>4797 ms (2342)</td>
</tr>
<tr>
<td>Other</td>
<td>‘Appropriate’</td>
<td>5398 ms (3155)</td>
<td>5569 ms (3358)</td>
</tr>
<tr>
<td></td>
<td>‘Inappropriate’</td>
<td>6227 ms (4756)</td>
<td>6015 ms (3060)</td>
</tr>
</tbody>
</table>

Note: Values in parentheses are standard deviations of the mean. All values are calculated from subjects included in RT analyses only.