Psihologijske teme, 28 (2019), 3, 483-506
Izvorni znanstveni rad UDK: 17.022
doi:https://doi.org/10.31820/pt.28.3.2

Deontic Moral Reasoning Task: Is Moral Reasoning Special?

Mislav Sudić, Pavle Valerjev
University of Zadar, Department of Psychology, Zadar, Croatia
Josip Ćirić
University of Zadar, Department of Information Sciences, Zadar, Croatia

Abstract

Domain theory suggests that moral rules and conventions are perceived differently and elicit a different response. A special procedure was designed to test this hypothesis in a laboratory setting using a deontic reasoning task. The goal was to gain insight into the cognitive and metacognitive processes of deontic reasoning from simple deontic premises. In the 3x2x2 within-subjects design, we varied rule-content (moral, conventional, abstract), rule-type (obligation, permission) and the induced dilemma (punishment dilemma, reward dilemma). Participants (N = 78) were presented with 12 laws. After memorizing a law, eight cases were presented to participants so that they make a quick judgment. Participants were tasked with punishing rule-violators, ignoring rule-conformists, and rewarding rule-supererogation. Response times (RT) and accuracy were measured for each judgment, and final confidence was measured after a set of judgments. No differences were expected between rule-types, except for superior performance for moral content and punishment dilemmas. RT correlated negatively with confidence levels, while accuracy correlated positively. Moral reasoning was more accurate than conventional and abstract reasoning, and produced higher confidence levels. Better performance was found for punishment dilemmas than reward dilemmas, likely due to the presence of a cheater-detection module; but the differences were not found in moral reasoning. Moral reasoning was also independent of rule-type, while conventional and abstract reasoning produced superior performance in obligation-type than in permission-type rules. A large drop-off in accuracy was detected for rules that allowed undesirable behaviour, a phenomenon we termed the "deontic blind spot". However, this blind spot was not present in moral reasoning. Three lines of evidence indicate a qualitative difference between the moral and other deontic domains: (1) performance for moral content was independent of rule-type, (2) moral content produced an equal activation of violator- and altruist-detection modules, and (3) moral content produces higher levels of confidence.

Keywords: moral reasoning, convention, metacognition, deontic logic

✉ Mislav Sudić, Department of Psychology, University of Zadar, Obala kralja Petra Krešimira IV 2, 23000 Zadar, Croatia. E-mail: *mislav.sudic@gmail.com*

Introduction

Prosocial behaviour was often an adaptive strategy to our ancestors (Dawkins, 1976), evolving a variety of moral cognitive tools (Greene, 2014). Cultural forces later sorted those tools into more coherent sets of ideas, rituals or philosophies (Haidt, 2012).

From the early 20th century, psychologists began to take an interest in what makes people act nobly and unselfishly (Haidt, 2008). Later, Kohlberg's (1976) research pioneered the field of modern moral psychology. His rationalist approach posited that through exploring and navigating their social environments, children develop their reasoning abilities in six stages of morality.

Moral and Social Domains

Turiel, Killen, and Helwig (1987) distinguish between the domain of morality and social convention. According to their definition, morality is concerned with topics of justice, rights and harm, while conventions (the social domain) are usually culturally determined and often arbitrary. Tisak and Turiel (1988) have found that children hold different opinions about transgressions in the two domains, and Blair (1997) found that children with psychopathic tendencies have difficulties in recognizing the moral/conventional distinction. Therefore, the main difference between the two domains seems to be that the moral input primes an affective response that adds more weight to the importance of moral rules.

The view that conventions should be considered as separate from morality has been criticized due to different cultures perceiving some of the conventional questions as a part of their societies' moral structure (Haidt, 2012). Nevertheless, the moral and conventional content for this study was selected using this distinction, since WEIRD samples (see Henrich, Heine, & Norenzayan, 2010), like the sample in this study, seem to conform better to the predictions of Domain Theory (Haidt, 2012).

Intuitionism

Recently, the role of strategic reasoning as the bedrock of moral reasoning has been subordinated in favour of the role of moral emotions and intuition, culminating in the formulation of the Social-intuitionist model (Haidt, 2001). It posits that a morally salient input causes an automatic emotional response that results in guiding the moral reasoning process. Strategic reasoning, from that point, is a source of justification and social propaganda.

So, how is the moral content of a stimulus recognized before the onset of strategic reasoning? And how does the automatic moral response interfere with the reasoning process? The answer to both questions is offered by the Social intuitionist's

sister-theory – the Moral Foundations Theory (Haidt, 2012). People differ in the sensitivity of six mechanisms: harm/care, fairness/cheating, loyalty/betrayal, authority/subversion, sanctity/degradation (Graham et al., 2011), and liberty/oppression (Iyer, Spassena, Graham, & Haidt, 2012). When the content of a stimulus is related to one of those foundations, then the more sensitive a person's foundation is, the more likely it is that a moralistic response will be triggered. This response influences moral judgment and drives the strategic process of justification. Haidt (2007) presents four lines of evidence for the Social intuitionist model, the most important of them being *moral dumbfounding*. First, a person is presented with a moral dilemma that creates a strong irrational response. Next, he is confronted by the interviewer with arguments that falsify the emotion-driven intuition. After exhausting all the reasons, the person stubbornly persists in the initial position, while admitting to being dumbfounded by his inability to articulate any rational arguments for that position (Haidt, 2001).

The moral content of rules in the current study was chosen to elicit a response from only the first two foundations: care/harm, and fairness/cheating. Although one of the premises of Social intuitionism is that *questions of justice and care do not cover the entire moral domain* (Haidt, 2001, 2007, 2008, 2012), these two foundations are endorsed by both liberals and (to a lesser degree) conservatives, while the rest are usually perceived as conventions by liberals. Since students in social sciences generally (Haidt, 2012), and at the University of Zadar in particular (Sudić & Didović, 2018) lean toward the left, we assumed they would react more strongly, on average, to the first two foundations – likely only recognizing them as part of the moral domain.

Metacognition and Dual-Processing

The divide between "rationalists" and "social intuitionists" is still alive, and some alternative theories of moral processing have also been proposed (e.g. Bucciarelli, Khemlani, & Johnson-Laird, 2008; Mikhail, 2007). One of them is Greene's (2014) dual-processing paradigm – an approach that seems to bridge that divide.

Greene (2014) claims that in everyday circumstances, we use simple heuristics (intuitions) to determine right from wrong in a way Social intuitionism predicts. These typically take the form of a deontological judgment (morality based on inner principles). However, when needed, or if properly cued, we can engage in deliberate reasoning that can override the initial intuition, and produce a utilitarian response (moral cost/benefit analysis), which is more in line with the Rationalist model. Most of the research within this paradigm was conducted using a variation of the trolley scenario: "a trolley is about to kill five people, but there is an option of redirecting it to only kill one person." A *deontological* response would be to not intervene (redirecting the trolley would be murder, which is intrinsically wrong), and a

utilitarian response would be to sacrifice the one person (killing one person to save five is a net-good).

In parallel, a different dual-processing paradigm was being developed within the reasoning literature (e.g. Ackerman & Thompson, 2015; Evans & Stanovich, 2013). The term "reasoning" is used differently from the one typically used in moral psychology. Thus far, we used it to denote *strategic and deliberate thinking* in moral dilemmas. In the cognitive literature, reasoning is a broader term that refers to *cognitive processes of drawing a conclusion from premises* (Kellogg, 1995), whether those premises are explicit or not (Bucciarelli et al., 2008). The "reasoning" task used in this study also refers to the latter definition.

According to the dual-process approach, cognitive processes can be divided into Type 1 and Type 2 thinking, the former being fast, automatic and based on heuristics; while the latter is conscious, logical and cognitively taxing (see Kahneman, 2013). Recently, Evans and Stanovich (2013) pointed out that the only remaining defining features of the two systems are *automaticity* for Type 1, and *cognitive decoupling* for Type 2 processes. The rest of these features simply happen to co-occur (e.g. Type 1 being fast, or Type 2 being conscious), but are non-essential properties.

Metacognition is a system that monitors these subsystems. It responds to subtle cues like *fluency* (the ease of response production), in order to mediate between the processes of the two systems (Thompson, 2009), while accuracy may or may not be tracked by metacognition (Shynkaruk & Thompson, 2006). It seems that in order for accuracy to influence metacognition, a response conflict has to actually be detected, i.e. a person must realize the difficulty of a task. Since the difficulty of a task will be proportional to its accuracy rate, if the conflict is detected, the metacognitive judgment should correlate with accuracy.

One can gauge the current status of the metacognitive system in multiple ways, for example by asking the participants to assess their level of confidence in a given answer (confidence judgment) or a set of given answers (final confidence judgment; Ackerman & Thompson, 2015). Final confidence was used in this study to gain insight into the higher cognitive processes after a set of judgments. We wanted to know whether people were more confident while reasoning under the influence of the moral affect, as is the case when reasoning from intuition, i.e. Type 1 (De Neys & Bialek, 2017). We expect that response time (a measure of fluency), but not necessarily accuracy, will predict confidence levels (Ackerman & Thompson, 2015).

Deontic Logic as Normative Framework

In order to obtain a measure of "accuracy", it is important to select its normative framework. In this case, normative accuracy will be defined as logical consistency in reasoning (in the text, the word "normative" will be dropped for brevity's sake). Deontic logic uses two basic deontic operators: *obligation* and *permission*. Since either can be negated, this creates four possible deontic categories: (1) obligation, (2)

non-obligation, (3) permission, and (4) non-permission. Act-individuals – actors for whom a given rule applies – have two possible action-values: performing and not-performing an action. This creates eight possible combinations of rules and action-values of act-individuals. These deontic relations fall into one of three categories: violation, conformity/indifference, and supererogation (Beller, 2010; Broersen et al., 2013; Heyd, 2016; Von Wright, 1951). Each of these relations demands an appropriate response in this experiment: punishment, ignoring, or reward, respectively (see Figure 1).

We only selected actions that were clearly either *desirable* or *undesirable*. Although formal deontic logic is neutral to the desirability of an action (Von Wright, 1951), in real life, almost without exception, obligations are used to govern desirable behaviours, while permissions govern undesirable ones. To illustrate this, consider the absurdity of a rule that permits desirable behaviour (e.g. you are permitted to donate to charity), the needlessness of a rule that does not obligate undesirable behaviour (e.g. you are not obligated to eat your child), or an irrationality of any rule that obligates one to do bad, or does not permit one to do good. In order to avoid confusion from the unrealistic structure of rules, desirable content was restricted to (non)obligation rules and the undesirable to (non)permission rules.

If one correctly memorizes a rule, after observing an act-individual, one of two possible dilemmas is induced: a punishment dilemma (to punish or ignore?) or a reward dilemma (to reward or ignore?). These dilemmas fit the parameters for the activation of cheater-detection (Cosmides & Tooby, 2015) and altruist-detection (Oda, Hiraishi, & Matsumoto-Oda, 2006) modules. A punishment dilemma is induced by rules that put an actual constraint on the action by either obligating or not permitting it. This means an act-individual can either conform to the rule or violate it. On the other hand, a reward dilemma is induced by rules that remove constraints on actions (non-obligations/permissions), thus providing liberty to act as one pleases. This gives the actor an option to conform to a lack of constraints, or freely choose the supererogatory (altruist) option.

Detection of Altruists and Cheaters

Using the deontic framework we can induce different types of dilemmas in participants by presenting them with a task to detect violators, conformists and the supererogatory. We expect that in order to easily solve one of these dilemmas, participants will engage one or more of the domain-specific reasoning algorithms. Since both the ability to detect cheaters (Cosmides & Tooby, 2015) and a concurrent ability to detect altruists (Oda et al., 2006) evolved to solve similar dilemmas (e.g. social exchange), it seems those two algorithms are the most likely to be used when solving a punishment or a reward dilemma, respectively. The endgame of altruist- and cheater-detection systems is to direct adaptive *action* (e.g. punishing violators

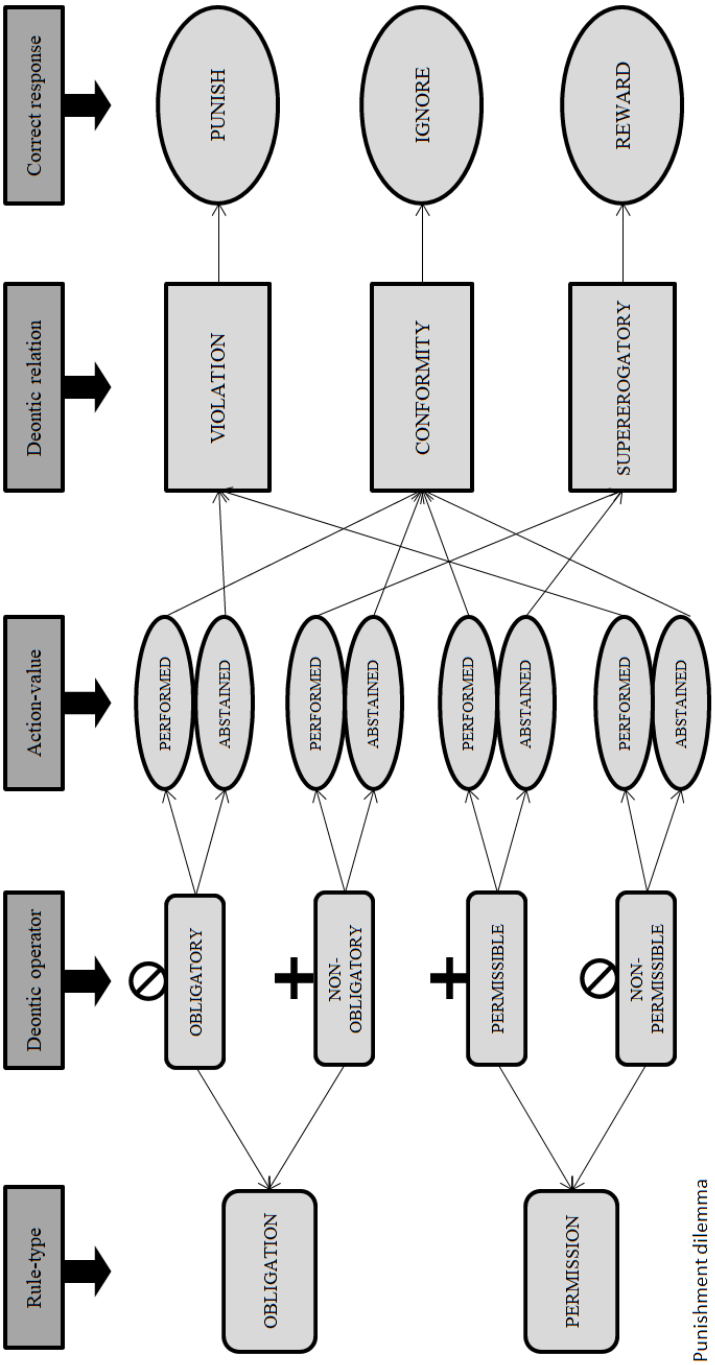

Figure 1. A representation of how different deontic operators (in rules) and action values (of act-individuals) form deontic relations (violation, conformity, or supererogatory), and a correct response (punishment, ignoring, rewarding).

and rewarding altruists), therefore participants in this study were not instructed to simply identify whether the act-individual was a violator, conformist or an altruist (a cognitive component), but rather to punish, ignore or reward them (behavioural reaction). We assumed that this will produce a more natural response from a participant than if he/she were simply given a logical reasoning task, thus increasing the external validity of the experiment without compromising its internal validity.

An inability to detect a violator was more likely to produce a catastrophic outcome for our ancestors rather than an inability to detect helpful individuals. While both systems offer an adaptive advantage to an individual, detecting cheaters is more difficult since they benefit from hiding their nefariousness. This puts adaptive pressure on the evolution of violation-detection sensitivity (Cosmides & Tooby, 2015). On the other hand, altruists tend to benefit from virtue signalling – making them easier to detect (Dawkins, 1976; Haidt, 2012), which likely puts relatively less of an adaptive pressure to evolve equal levels of altruist detection sensitivity. We thus hypothesized that the cheater detection algorithm will produce faster and possibly more accurate results.

Deontic Reasoning Task

According to Beller (2010), the most popular way to measure deontic reasoning is using deontic forms of Wason's task (e.g. Kellogg, 1995; Oda et al., 2006). However, the Wason's task is limited to conditional (if-then) reasoning.
Instead, we constructed a task that resembles a syllogism:

Major (deontic) premise: It's (not) obligatory/permissible to do A.

Minor premise: A person is (not) doing A.

Conclusion: Therefore, the person is a violator/conformist/supererogatory.

Deontic tasks commonly ask a participant to identify a correct answer or identify an activity (e.g. Oda et al., 2006; see Beller, 2010). We went further and measured a form of deontic behaviour: participants were tasked with not just sorting out violators, conformists and altruists. They were tasked with appropriately punishing, ignoring and rewarding them. Both response times and accuracy were recorded in this task, as well as final confidence judgments.

In summation, the goal of this study is to determine how moral, conventional and abstract rules were processed based on their deontic type and the dilemma they induce. We expected the best performance in moral rules and better performance in conventional than in abstract rules. Judgments of confidence were expected to follow the same trend. We did not expect to find a difference in obligation or permission type rules but expected a faster and more accurate response to punishment than reward dilemmas. We expected to find a negative correlation between confidence and response time but not necessarily a correlation between accuracy and confidence. However, if the response conflict was in fact detected, then the confidence-accuracy correlation is expected to be positive.

Method

Participants

The sample (N = 78, 67 female) was recruited among students of psychology at the University of Zadar, ages 19-28 (M = 21.62, SD = 2.07). Participants were recruited using the convenience method through social media.

Design

The design of the experiment was 3x2x2 within groups. The independent variables were:

a) Content: moral, conventional, or abstract
b) Rule type: obligation, or permission
c) Induced dilemma: punishment dilemma (to punish or to ignore), or reward dilemma (to reward or to ignore)

The dependent variables were response time, accuracy and final confidence.

Materials

The independent variables were manipulated through the deontic reasoning task form (rule type and induced dilemma) and content.

Rule type and induced dilemma. Tasks were created based on the type of operator used in rules. Obligations provide information on (non-)obligatory actions, thus governing desirable behaviour, while permissions govern whether undesirable actions are permissible. This is an improvement on the typical form of a deontic task (see Beller, 2010), due to the fact that banning or allowing desirable behaviour, as well as (non-)obligating undesirable behaviour is either absurd or unnecessary in real life. Therefore, experimental stimuli are not wasted on nonsensical rules. "Soft" operators (non-obligation, and permission) are designed to induce reward dilemmas (to reward or to ignore), while "hard" operators (obligation, and non-permission) induce punishment dilemmas (to punish or ignore).

Content. The content was selected by loosely referring to Haidt's Moral foundations theory and Turiel's Domain theory. Moral content was selected using the first two foundations, e.g. the law governing deception and honesty. Conventional content was selected by referring to Turiel's morality-convention distinction, e.g. the law governing car parking. Abstract content was selected by replacing a name of an action with an uppercase letter, e.g. "It's obligatory to perform A". See Appendix A for a detailed selection of the laws. Within every law, one rule was designed to induce the reward dilemma, the other to induce the punishment dilemma.

Accuracy. Participants were tasked with determining the relationship between the prescription of the law and the behaviour of a person. There are three deontic relations to the rules: violation, conformity, or supererogatory. Participants had three options: ignoring, punishing, or rewarding. It is considered accurate to:

a) ignore those that conform to rules
b) punish violators (those that *do not perform* desirable obligatory acts, or *perform* undesirable impermissible acts)
c) reward the supererogatory (those that *perform* desirably even when not obligated, or *do not perform* undesirable acts even when they are permissible)

Condition balancing. Participants had three possible reactions at their disposal in the experiment (P-punishing, R-rewarding, I-ignoring), and they reacted to them with the index, middle and ring finger of their right hand. It was determined in four preliminary studies that simple reaction times significantly differed between the three fingers. Therefore, participants were rotated through three conditions of finger-to-key combinations: IPR, PRI and RIP (the first letter represents the index finger, the second the middle finger, the third the ring finger). The IPR condition, for example, means that the index finger was matched with the reaction of ignoring, the middle finger with punishing, and the ring finger with reward. Secondly, participants were also rotated in two other conditions of rule sequencing. For example, if those in sequence "a" had a rule ordering within one law of XY, those in sequence "b" had the reverse, YX ordering – so that primacy effects could be controlled.

Procedure

The experiment was conducted in the Laboratory for Experimental Psychology at the University of Zadar. Participants sat across the computer screen, and had a written part of the instructions in front of them. The experiment was designed using E-Prime version 2.0.10.356. A paper with feedback questions was provided after completing the experiment. It included questions about demographics (sex, age, subject of study), and a question whether they understood the instructions.

The procedure consisted of three phases: (1) two practice tasks, (2) the main task, (3) feedback. Before the procedure, participants were asked to review instructions on the task manual in front of them. They were asked if they had any questions, and informed that they would not be allowed to ask anything further once the task starts.

Practice tasks. The first practice task was designed to accustom the participant to reacting with the index finger, middle finger, and the ring finger of the right hand. It consisted of 30 trials. The second practice task was similar, but instead of reacting to numbers corresponding to fingers, participants practiced pairing up fingers with

reactions of ignoring, punishing, or rewarding – depending on conditions assigned to them. The practice consisted of 90 trials.

Deontic reasoning task. The DRT consisted of 12 major tasks (each corresponding to one law), and 8 minor tasks within each of the major tasks. The sequence of major tasks was randomized, as was the sequence of minor tasks within them. When one of the major tasks began, the participants were presented with a law, a short description of the law, and two rules of which the law consisted. The instructions for the task were provided in great detail, including task structure, how to react, and what is considered the correct answer. They were instructed to take as much time as they needed to memorize the law. Once the participant signaled the program that he/she memorized the law, he/she was presented with eight cases (eg. *John didn't shoplift.*) in random order. Four cases referred to the first rule, four to the second. Half of those cases performed the action, half abstained from action. Half of names were female, the other half male. As soon as the participant reacted by rendering judgment, he/she was presented immediately with the next case until all eight cases (minor tasks) were exhausted. Response times and accuracy were recorded during the rendering of the judgment. After the minor tasks ended, participants were asked to rate their final confidence on a scale of 1-7 with the question: *How confident are you of your performance in the previous task?* See Table 1.

Feedback. After completion, participants were asked to fill out a feedback questionnaire, where we asked them whether they understood the performed task, and collected data about age and gender.

Table 1

Examples of Steps of Deontic Tasks Based on the Three Contents

	Morality	**Convention**	**Abstract**
Step 1: Memorize a law with two rules (unlimited time).	1. It's not obligatory to tell the truth. $^{TO\text{-}DR}$ 2. It's not permissible to betray a secret. $^{TP\text{-}DP}$	1. It's obligatory to dress formally. $^{TO\text{-}DP}$ 2. It's not obligatory to compliment the chef. $^{TO\text{-}DR}$	1. It's obligatory to do A. $^{TO\text{-}DP}$ 2. It's permissible to do B. $^{TP\text{-}DR}$
Step 2: Punish, reward or ignore a total of eight people (time-sensitive).	Sam did tell the truth. (accurate: reward).	Elliot didn't compliment the chef. (accurate: ignore)	Jonathan didn't do A. (accurate: punish)
Step 3: Final confidence judgments	How confident are you of your performance in the previous task? (not confident) **1 – 2 – 3 – 4 – 5 – 6 – 7** (very confident)		

^{TO} = obligation type rule; ^{TP} = permission type rule; ^{DP} = punishment dilemma; ^{DR} = reward dilemma

Results

In total, there were 96 judgments (8 reactions within each of 12 laws) rendered by a participant, and response times and accuracy were recorded and contrasted against content, type and induced dilemma. Since there were eight judgments per situation, median values of response times and accuracy were calculated for each situation. All except one experimental situation for response time, and two for accuracy, were within an acceptable range of +/-2 in measures of skewness and kurtosis. After excluding all results from those results that deviated more than 2.5 standard deviations from the mean, skewness and kurtosis was reduced to a +/-1 range. One participant was excluded for being too aberrant (+/- 3 standard deviations from the mean in multiple variables), and two for reporting that they did not understand the instructions. For descriptive data see Table 2 and Appendix B.

Table 2

Descriptive Statistics for Main Effect of Content on Confidence, and Three-Way Interaction Effects of Content, Rule Type and Dilemma on Response Times and Accuracy

			Content					
			Abstract		Convention		Morality	
			M	*SD*	*M*	*SD*	*M*	*SD*
	Rule type	Dilemma						
Response time (ms)								
	Obligation	Punishment	2482	730	2187	658	2534	736
		Reward	2984	1014	2451	887	2777	1062
	Permission	Punishment	3268	1171	3157	1089	2962	1192
		Reward	3301	1276	3243	1171	2912	973
Level of accuracy								
	Obligation	Punishment	.87	.16	.89	.17	.83	.21
		Reward	.83	.19	.90	.15	.88	.15
	Permission	Punishment	.79	.20	.78	.18	.83	.18
		Reward	.58	.32	.62	.23	.86	.16
Confidence								
			4.30	1.26	4.41	1.28	4.66	1.33

Response Time and Accuracy

A three-way (3x2x2) within-subjects ANOVA was calculated for both response time and accuracy. All three main effects – content, rule-type, and dilemma - were significant for both dependent variables. Furthermore, all two-way interaction effects were significant for accuracy. However, only content x type, and type x dilemma interactions were found for response times (see Table 3). Post-hoc Bonferroni tests for the main effect of content, as well as for all three two-way interactions were

performed for both response time and accuracy with p = .05 as the significance threshold.

Table 3

ANOVA Results for Response Time and Accuracy Depending on Content, Rule Type and Dilemma

	Response times			Accuracy		
	F	*df*	Partial η^2	*F*	*df*	Partial η^2
Content	11.33**	2/132	.146	19.71**	2/134	.227
Rule	97.35**	1/ 66	.596	121.47**	1/ 67	.645
Dilemma	12.26**	1/ 66	.157	13.47**	1/ 67	.167
Content x Rule	13.79**	2/132	.173	32.31**	2/134	.325
Content x Dilemma	2.83	2/132	.041	19.85**	2/134	.229
Rule x Dilemma	5.34*	1/ 66	.075	19.27**	1/ 67	.223
Content x Rule x Dilemma	1.91	2/132	.028	4.67*	2/134	.065

* $p < .05$; ** $p < .01$.

Main effects. Participants performed better in moral than abstract reasoning. Conventional reasoning was slower than moral, but more accurate than abstract reasoning. They performed better in obligation reasoning than permission reasoning, a result that produced a large effect size (RT: η_p^2 = .596; accuracy: η_p^2 = .645). Participants were also better in solving punishment dilemmas than reward dilemmas.

Content x Rule type. Within all contents, participants performed faster while reasoning with obligations, though the gap seems to be narrower in moral content, and widest in conventional. A similar pattern is seen for accuracy: the gap was very wide for the conventional and abstract content, but it disappeared when it came to moral reasoning. There was no difference in accuracy of obligation reasoning across different contents. However, permission reasoning was lower in cases of abstract and conventional content, but significantly higher in moral content. Participants produced the fastest response for obligations in the conventional domain, but they were fastest to react to moral content when faced with permissions.

Content x Dilemma. Punishment dilemmas were solved with identical success across rule contents, but reward dilemma accuracy climbed on a linear slope upwards from abstract to conventional to moral content, where it converged with punishment dilemma accuracy (Figure 2, right graph). Punishment dilemmas produced a faster response in conventional than in abstract reasoning. On the other hand, reward dilemmas produced a slower response in abstract reasoning as opposed to conventional and moral reasoning (Figure 2, left graph). Performance for solving punishment and reward dilemmas was equal in moral reasoning, but they were faster and more accurate in solving the punishment dilemma when reasoning abstractly, and more accurate during conventional reasoning.

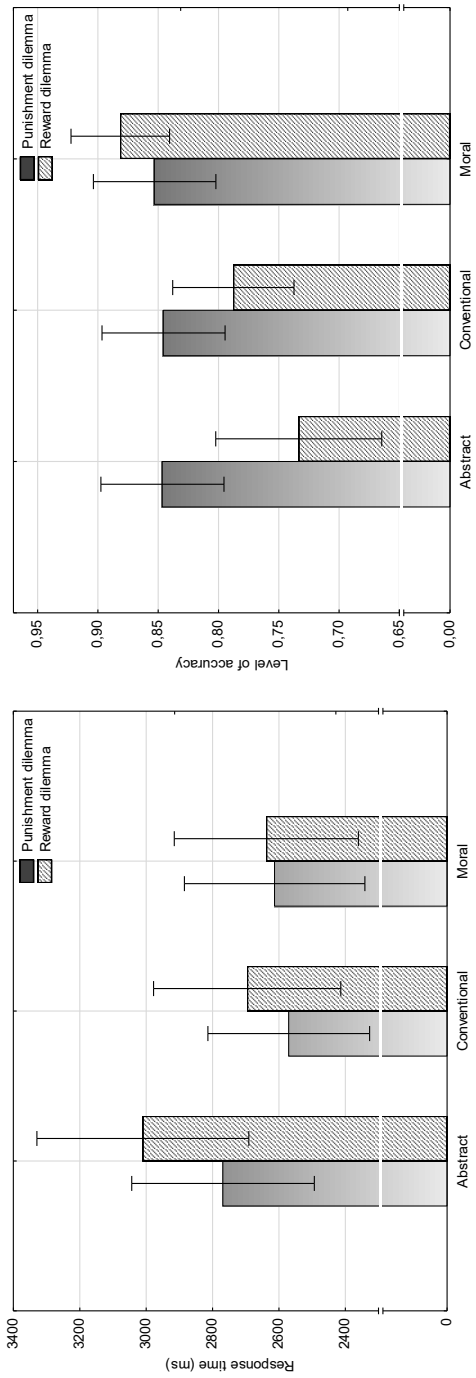

Figure 2. Response time (left) and level of accuracy (right) as a function of content and dilemma (spreads represent 95% confidence).

Rule type x Dilemma. Keeping the content constant, participants solved the punishment dilemmas faster when reasoning about obligations, and more accurately in permission-type rules. While solving both dilemmas, participants were faster and more accurate in obligations than permissions.

A deontic blind spot. As is obvious from Appendix C (right graph), there is a sharp decline of 21-25% in accuracy during tasks where undesirable actions were permitted (permission x reward dilemma situation). We termed this phenomenon a "deontic blind spot", and will expand upon this in the Discussion. However, it only affected conventional and abstract contents. Moral content produced a compensatory effect for the decline – accuracy did not differ between the four type x dilemma experimental situations.

Judgments of Confidence

Confidence levels were measured for three different contents: abstract, conventional and moral. Two types of analyses were performed to determine: (1) does confidence differ between contents?, and (2) does confidence correlate to response times or accuracy? Results for final confidence, as well as those for RT and accuracy, were first averaged between moral, conventional and abstract rules for every participant.

Main effect. We performed a one-way within-subjects ANOVA on confidence levels between three contents. A significant content effect was found ($F(2, 154) = 5.31$; $p < .01$; $\eta_p^2 = .068$), and post-hoc analysis pointed to higher confidence in moral than in abstract content (for descriptive data see Table 2).

RT and accuracy as predictors. In order to determine whether reaction time and accuracy separately predict the level of confidence, three multiple regression analyses were performed with confidence as the criterion, and response time and accuracy as predictors. Each analysis was conducted for each of the three contents. All three analyses found a significant correlation ($R^2(2/75)$ .22 to .26, $p < .01$). Confidence was predicted negatively by response time ($\beta = -.28$ to $-.39$, $p < .01$), and positively by accuracy ($\beta = .24$ to .35, $p < .05$).

Discussion

The goal of the study was to determine how different contents of rules, based on their deontic type and the dilemmas they induce, produce different performance and confidence outcomes. Content was varied between moral, conventional and abstract rules. The deontic type and induced dilemma depended on the deontic operator in the rule: *obligation* (obligation-type, punishment dilemma), *non-obligation* (obligation-type, reward dilemma), *permission* (permission-type, reward dilemma), and *non-permission* (permission-type, reward dilemma).

Content Effects

In abstract deontic reasoning participants were on average significantly faster by 450-759 ms in processing rules that used the deontic operator *obligatory* (obligation rule type, punishment dilemma), than rules with the other three types of operators. However, they were also 20-27% significantly less accurate when reasoning with rules that used the operator *permissible* (Appendix C, right graph).

It seems that abstract rules using the *obligation* operator produced a more fluent response, while *permission* operators produced a less accurate one. *Non-obligation* and *non-permission* operators produced a response that did not significantly differ from *permission* operators in response time, and with *obligation* operators in accuracy. A similar pattern was found when using conventional content. The difference is, performance with conventional rules seemed to be more dependent on the type of rule, while performance with abstract rules was more dependent on the induced dilemma. On the other hand, moral content produces a more uniform response across the other two variables. It also produced higher levels of confidence.

As we hypothesized, response time correlated negatively with confidence ratings in all three contents. Interestingly, accuracy was found to correlate positively with confidence, a less common finding in the reasoning literature (see metacognition and dual processing in the Introduction), indicating response conflict was, for some reason, detected by metacognition (Ackerman & Thompson, 2015; Shynkaruk & Thompson, 2006). Furthermore, lower confidence for abstract rules than moral rules indicated that System 2 was more likely to be utilized in the abstract task, while moral rules were solved more intuitively (De Neys & Bialek, 2017) – which is in line with the Social Intuitionist Model (Haidt, 2012).

Obligations and Permissions

Although deontic logic permits any kind of content combined with any kind of deontic operator (Von Wright, 1951), in real life obligations govern desirable actions, and permissions govern undesirable ones. Therefore, it is impossible to disassociate the desirability of an action from the rule-type in this study.

Results indicate better performance for obligation-type than for permission-type rules, and the effect sizes were considerable. Two different effects seem to simultaneously contribute: (1) an increase in performance when reasoning with conventional obligations, and (2) a "blind spot" for reasoning about rules that permit undesirable behaviour. The first effect is likely an artifact of convenient sampling – students might have more experience with reasoning around conventional obligations, leading to a more nuanced schema (e.g. do I *have to* do this homework, or is there a way around it?), while permission-type conventions tend to be more imperative in nature, thus possibly perceived as less malleable. Of course, this is purely speculative, and should be examined in subsequent studies. However, the fact

remains that the fastest and most accurate responses in the study were to conventional obligation-type rules. The "blind spot" effect will be discussed in more detail later.

Rule-Violation Bias

Every task induces only one of two types of dilemmas: either *a punishment dilemma* (to punish or to ignore) or *a reward dilemma* (to reward or to ignore). The dilemmas are induced by the type of constraint the rule imposes. Evolutionary and game theoretical modeling (Cosmides & Tooby, 2015) posits a set of evolved domain-specific modules that enable adaptive reasoning. An evolved cheater-detection module makes people more biased toward the detection of rule violations, correctly predicting better performance in punishment dilemmas in the case of abstract and conventional content. This is perfectly illustrated in Figure 2 (right graph), where the accuracy rate for punishment, while controlling rule-type, is stable at 85%, independent of content.

However, moral rules are an exception. While the ability to correctly solve the punishment dilemma remains constant across different contents, both the response times and accuracy for the two types of dilemmas seem to converge when reasoning with moral rules. Therefore, the activation of the altruist-detection module seems to match violation-detection only if the moral affect has been previously primed.

It should be noted that sorting out violators, conformists, and altruists is not the same as reacting to them – even though participants were provided with precise instructions on how to solve the given tasks properly. Someone might be perceived as an altruist, or a violator, but if a participant considers the scope of the supererogatory act, or the severity of the violation as mild, he/she might not opt for a reward or punishment. Subsequent studies may examine in what degree this might be the case.

A Deontic Blind Spot

There seems to be a sharp drop in accuracy in situations that allow (undesirable) actions (see Figure 3). In such situations, a permission-type rule implies undesirable action, something that has been explicitly pointed out to participants in the instructions. Although the action is undesirable, since it is permitted, it should induce a reward dilemma. However, this does not often occur, a phenomenon we will refer to here as the "deontic blind spot". For example:

It's *permissible* to do M.

Carol did M.

Melinda did *not* M.

Should Carol be: punished, ignored, or rewarded?

Should Melinda be: punished, ignored, or rewarded?

Correct answers: Carol should be ignored, and Melinda should be rewarded. An average accuracy in this situation was about 61%. Curiously, this may not be accounted for entirely due to a lack of concrete content. For example:

It's *permissible* to use your credit card.

Carl used his credit card.

Tim didn't use his credit card.

Should Carl be: punished, ignored, or rewarded?

Should Tim be: punished, ignored, or rewarded?

Correct responses were that Carl should be ignored, and Tim rewarded. Here, conventional content was used, yet the accuracy only went up (non-significantly) to 65%. What makes finding an explanation for this phenomenon more difficult is the fact that when the rule is framed in moral terms, the "blind spot" seems to disappear, and the accuracy is approximately 86%.

Any explanation for this phenomenon must include answers to the questions (1) *why is there a drop off in accuracy?,* and simultaneously (2) *why does moral content compensate for it?* The omission bias (underestimating the importance of harmful avoidance) might explain why accuracy dropped (DeScioli, Bruening, & Kurzban, 2011), but it would not explain why this effect was not present in moral rules. One possibility is that moral reasoning might be qualitatively different from other forms of deontic reasoning. However, this still leaves us with the challenge of explaining the existence of the deontic blind spot in conventional and abstract reasoning.

Is Moral Reasoning Special?

In order to gain insight into moral reasoning processes, we contrasted the performance for rules with the moral content with two control contents: abstract and conventional. Abstract content was used to control the concrete content effect. Social conventions differ from moral rules because they lack intrinsic value. As the social-intuitionist model predicts, moral content produces an automatic affect, thus likely priming a different deontic schema than abstract or conventional content.

The existence of a special "moral reasoning schema", at least for reasoning with simple deontic premises, is supported by three findings in this study:

(1) Moral, unlike conventional or abstract, deontic reasoning, does not depend on whether the rule is an obligation that governs desirable action, or a permission that governs undesirable ones.

(2) Moral reasoning does not seem to favor violator-detection over altruist-detection.

(3) Moral reasoning is followed by higher confidence.

This is not to say moral reasoning is not a type of deontic reasoning, rather it seems to engage different cognitive processes. Of course, many important factors

have not been controlled in this limited study, so that conclusion might in time prove to be premature.

Future studies should focus on eliminating insufficiencies of this study: better practice tasks, empirically selected and balanced rules (e.g. we are unsure to what degree conventional and abstract rules were saturated with moral content), confidence judgments after every reaction, simpler instructions, etc. Also, to find the explanation for the "deontic blind spot" phenomenon, which we failed to account for. The task used here can be modified for a variety of research problems, and using deontic logic as a moral reasoning normative in general may help to provide better insight into our moral reasoning abilities. However, this approach is limited because the "correct" response in this study is only defined as logical consistency.

Conclusions

Results indicate that: (1) moral rules are easier to process than conventional ones, and conventional rules easier than abstract ones, (2) punishment dilemmas are easier to solve than reward dilemmas, (3) obligations are easier to process than permissions, (4) confidence is higher for moral than abstract rules, and (5) in all three contents of rules it correlates negatively with response time, and positively with accuracy.

References

Ackerman, R., & Thompson, V. A. (2015). Meta-reasoning: What can we learn from meta-memory? In A. Feeney & V. A. Thompson (Eds.), *Reasoning as memory* (pp. 164-182). London and New York: Psychology Press. doi:10.1037/t11859-000

Beller, S. (2010). Deontic reasoning reviewed: Psychological questions, empirical findings, and current theories. *Cognitive Processing, 11*(2), 123-132. doi:10.1007/s10339-009-0265-z

Blair, R. J. R. (1997). Moral reasoning and the child with psychopathic tendencies. *Personality and Individual Differences, 22*(5), 731-739. doi:10.1016/S0191-8869(96)00249-8

Broersen, J., Gabbay, D., Andreas, H., Lorini, E., Meyer, J.-J., Parent, X., & van der Torre, L. (2013). Deontic logic. In S. Ossowski (Ed.), *Agreement technologies* (pp. 171-179). New York: Springer Publishing Company. doi:10.1007/978-94-007-5583-3

Bucciarelli, M., Khemlani, S., & Johnson-Laird, P. N. (2008). The psychology of moral reasoning. *Judgment and Decision Making, 3*(2), 121-139.

Cosmides, L., & Tooby, J. (2015). Adaptations for reasoning about social exchange. In D. M. Buss (Ed.), *The handbook of evolutionary psychology, Second edition. Volume 2: Integrations* (pp. 625-668). Hoboken, NJ: John Wiley & Sons.

Dawkins, R. (1976). *The selfish gene*. Oxford: Oxford University Press.

De Neys, W., & Bialek, M. (2017). Dual processes and conflict during moral and logical reasoning: A case for utilitarian intuitions? In B. Trémolière & J. F. Bonnefon (Eds.), *Moral inferences* (pp. 123-136). Hove, UK: Psychology Press.

DeScioli, P., Bruening, R., & Kurzban, R. (2011). The omission effect in moral cognition: Toward a functional explanation. *Evolution and Human Behaviour, 32*(3), 204-215. doi:10.1016/j.evolhumbehav.2011.01.003

Evans, J. St. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science, 8*(3), 223-241. doi:10.1177/1745691612460685

Graham, J., Nosek, B. A., Haidt, J., Iyer, R., Koleva, S., & Ditto, P. H. (2011). Mapping the moral domain. *Journal of Personality and Social Psychology, 101*(2), 366-385. doi:10.1037/a0021847

Greene, J. (2014). Beyond point-and-shoot morality: Why cognitive (neuro)science matters for ethics. *Ethics, 124*(4), 695-726. doi:10.1086/675875

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108*(4), 814-834. doi:l0.l037//n033-295X.108.4.814

Haidt, J. (2007). The new synthesis in moral psychology. *Science, 316*(5827), 998-1002. doi:10.1126/science.1137651

Haidt, J. (2008). Morality. *Perspectives on Psychological Science, 3*(1), 65-72. doi:10.1111/j.1745-6916.2008.00063.x

Haidt, J. (2012). *The righteous mind: Why good people are divided by religion and politics.* New York: Pantheon.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *The Behavioural and Brain Sciences, 33*(2-3), 61-135. doi:10.1017/S0140525X0999152X

Heyd, D. (2016). Supererogation. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. Retrieved from https://plato.standford.edu/

Iyer, R., Spassena K., Graham, J., & Haidt, J. (2012). Understanding libertarian morality: The psychological dispositions of self-identified libertarians. *PLoS ONE, 7*(8), 1-23. doi:10.1371/journal.pone.0042366

Kahneman, D. (2013). *Misliti, brzo i sporo [Thinking, fast and slow]*. Zagreb: Mozaik knjiga.

Kellogg, R. T. (1995). *Cognitive psychology*. London, UK: SAGE Publications.

Kohlberg, L. (1976). Moral stages and moralization: The cognitive-developmental approach. In T. Lickona (Ed.), *Moral development and behaviour: Theory, research, and social issues* (pp. 31-53). New York: Holt, Rinehart and Winston.

Mikhail, J. (2007). Universal moral grammar: Theory, evidence and the future. *Trends in Cognitive Sciences, 11*(4), 143-152. doi:10.1016/j.tics.2006.12.007

Oda, R., Hiraishi, K., & Matsumoto-Oda, A. (2006). Does an altruist-detection cognitive mechanism function independently of a cheater-detection cognitive mechanism? Studies using Wason selection tasks. *Evolution and Human Behaviour, 27*, 366-380. doi:10.1016/j.evolhumbehav.2006.03.002

Thompson, V. A. (2009). Dual process theories: A metacognitive perspective. In K. Frankish & J. St. B. T. Evans (Eds.), *In two minds: Dual processes and beyond* (pp. 171-195). Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780199230167.001.0001

Shynkaruk, J. M., & Thompson, V. A. (2006). Confidence and accuracy in deductive reasoning. *Memory & Cognition, 34*(3), 619-632. doi:10.3758/bf03193584

Sudić, M., & Didović, V. (2018). *Libertarijanci i moralno rasuđivanje [Libertarians and moral reasoning]*. Paper presented at XXI Psychology Days in Zadar conference, Zadar, Croatia. Retrieved from http://www.unizd.hr/Portals/29/2016/2018/XXI.%20Dani%20psihologije-%20Knjiga%20sa%C5%BEetaka.pdf?ver=2018-09-07-111105-703

Tisak, M. S., & Turiel, E. (1988). Variation in seriousness of transgression and children's moral and conventional concepts. *Developmental Psychology, 24*(3), 352-357. doi:10.1037/0012-1649.24.3.352

Turiel, E., Killen, M., & Helwig, C. C. (1987). Morality: Its structure, function, and vagaries. In J. Kagan & S. Lamb (Eds.), *The emergence of morality in young children* (pp. 155-243). Chicago, University of Chicago Press.

Von Wright, G. H. (1951). Deontic logic. *Mind, 60*(237), 1-15.

Zadatak moralnoga deontičkog rasuđivanja: Je li moralno rasuđivanje posebno?

Sažetak

Teorija domena pretpostavlja da se moralna i konvencionalna pravila drugačije percipiraju i rezultiraju različitim odgovorima. Osmišljena je procedura za testiranje ove hipoteze u laboratorijskim uvjetima koristeći zadatak deontičkog rasuđivanja. Cilj je bio dobiti uvid u kognitivne i metakognitivne procese deontičkog rasuđivanja polazeći od jednostavnih deontičkih premisa. Korištenjem nacrta 3x2x2 s ponovljenim mjerenjima manipulirali smo sadržajem pravila (moralna, konvencionalna, apstraktna), tipom pravila (obaveze, dopuštenja) i induciranom dilemom (dilema kažnjavanja, dilema nagrađivanja). Sudionicima (N = 78) prikazano je 12 zakona. Nakon što su zapamtili zakon, prezentirano im je osam slučajeva za koje su morali donijeti brzu odluku. Zadatak im je bio kažnjavanje prekršitelja, ignoriranje konformista i nagrađivanje supererogatornih. Mjereno je vrijeme odgovora i točnost za svaku odluku te konačna sigurnost nakon jednog niza odluka. Nisu očekivane razlike između tipova pravila, ali je očekivana bolja izvedba kod moralnih sadržaja i dilema kažnjavanja. Vrijeme je odgovora bilo negativno, a točnost pozitivno povezana s razinom sigurnosti. Moralno rasuđivanje bilo je točnije od konvencionalnog i apstraktnog te je dovelo do više razine sigurnosti. Bolja je izvedba utvrđena pri dilemama kažnjavanja u usporedbi s nagrađivanjem, vjerojatno zbog prisutnosti modula za detekciju varalica, ali te razlike nisu utvrđene pri moralnom rasuđivanju. Moralno je rasuđivanje također bilo neovisno o tipu pravila, dok su konvencionalno i apstraktno rasuđivanje doveli do bolje izvedbe pri obavezama nego dopuštenjima.

Velik je pad u točnosti utvrđen za pravila koja su dopuštala nepoželjna ponašanja, što je fenomen koji smo nazvali "deontička slijepa pjega". Ipak, ova slijepa pjega nije bila prisutna pri moralnom rasuđivanju. Zaključno, rezultati upućuju na kvalitativne razlike između moralne domene i ostalih: (1) izvedba pri moralnom sadržaju nije ovisila o tipu pravila, (2) moralni je sadržaj proizveo jednaku aktivaciju modula detekcije varalica i altruista te (3) moralni je sadržaj proizveo viši stupanj sigurnosti.

Ključne riječi: moralno rasuđivanje, konvencije, metakognicija, deontička logika

Primljeno: 27.10.2018.

APPENDIX A.

List of Laws and Rules Used in the Deontic Reasoning Task

CONTENT	LAWS/RULES
Abstract	LAW 123: It's obligatory to do A. It's permissible to do B. LAW 456: It's obligatory to do X. It's not obligatory to do Y. LAW 789: It's not obligatory to do R. It's not permissible to do T. LAW 000: It's permissible to do M. It's not permissible to do N.
Conventional	ID registration: It's obligatory to present proof of citizenship. It's permissible to pay later. Restaurant etiquette: It's obligatory to dress formally. It's not obligatory to compliment the chef. Car parking: It's not obligatory to turn on the blinkers. It's not permissible to take off your belt. Paying taxes: It's permissible to use your credit card. It's not permissible to use coins.
Moral	House pets: It's obligatory to feed your pets. It's permissible to abandon your pet. Violence and intervention: It's obligatory to perform first aid. It's not obligatory to report violence. Lying and honesty: It's not obligatory to tell the truth. It's not permissible to betray a secret. Store delinquency: It's permissible to overspend. It's not permissible to shoplift.

Note: Loosely translated from Croatian.

APPENDIX B.

Descriptive Statistics for Main Effects and Two-Way Interactions of Content, Rule-Type and Induced Dilemma

	Response time		Accuracy	
	M	*SD*	*M*	*SD*
Main effects				
Content				
Abstract	3062	914	.76	.17
Conventional	2834	889	.79	.15
Moral	2841	903	.84	.14
Rule-type				
Obligations	2583	695	.87	.13
Permissions	3160	903	.74	.15
Dilemma				
Punishment	2783	727	.83	.14
Reward	2958	846	.78	.14
Two-way interactions				
Content x Type				
Abstract Obligations	2734	777	.85	.15
Abstract Permissions	3305	1087	.69	.21
Conventional Obligations	2348	744	.89	.14
Conventional Permissions	3226	1052	.69	.18
Moral Obligations	2663	816	.86	.15
Moral Permissions	2957	999	.84	.15
Content x Dilemma				
Abstract Punishment dilemmas	2871	820	.83	.16
Abstract Reward dilemmas	3172	1065	.70	.21
Conventional Punishment dilemmas	2702	800	.83	.16
Conventional Reward dilemmas	2868	938	.76	.16
Moral Punishment dilemmas	2762	863	.83	.16
Moral Reward dilemmas	2845	921	.87	.14
Type x Dilemma				
Obligation Punishment dilemmas	2414	619	.86	.16
Obligation Reward dilemmas	2756	870	.87	.13
Permission Punishment dilemmas	3169	977	.80	.15
Permission Reward dilemmas	3162	944	.68	.20

APPENDIX C.

Response Times (Left Graph) and Accuracy (Right Graph) as a Function of Content, Rule Type, and Dilemma (Spreads Represent 95% Confidence)