PSIKOLOGIKA JURNAL PEMIKIRAN DAN PENELITIAN PSIKOLOGI Volume 30 Nomor 2, Juli 2025: 315-346
DOI:10.20885/psikologika.vol30.iss2.art9
E-ISSN: 2579-6518
P-ISSN: 1410-1289

Gender Difference in Responses to Moral Dilemmas: An Experimental Study on Dual Process Theory of Moral Judgment

Sabiqotul Husna
Psychology Study Program, Faculty of Social Science and Humanities, State Islamic University Sunan Kalijaga, Yogyakarta, Indonesia

Abstract. Morality is defined as the basis for evaluating acceptable and inacceptable behaviors. This study examined how gender and dilemma type influenced moral judgment, emotional arousal, and valence in 60 participants. Furthermore, an experimental method was adopted using a factorial design and a vignette-based scenario approach. The mixed factorial ANOVA analysis showed that moral decision had a significant effect on judgment, and deontological responses were rated more affirmatively than utilitarian. Deontological judgments prioritized adherence to moral rules or duties regardless of outcomes. Meanwhile, utilitarian judgments focused on the consequences of actions, aimed to maximize overall well-being. Gender did not significantly affect moral acceptability or decision type, with females reportedly exhibiting higher emotional arousal than males. Additionally, dilemma type significantly influenced both arousal and acceptability. Incidental harm was most acceptable, and prompted the highest arousal. Valence ratings were consistently negative across conditions, showing the distressing nature of moral conflict. These findings supported dual-process models of moral cognition, outlining the dominant role of emotional arousal and contextual factors over stable individual traits such as gender, and informed the development of ethics training and decision-making interventions in environments where professionals were often faced with morally complex decisions under pressure.

Keywords: deontological moral judgment, incidental moral dilemma, moral dilemma, utilitarian moral judgment, vignette experiment

Perbedaan Gender dalam Respon terhadap Dilema Moral: Sebuah Studi Eksperimental tentang Teori Proses Ganda Penilaian Moral

Abstrak. Moralitas selalu menjadi dasar dalam mengevaluasi perilaku manusia terkait apa yang dianggap dapat diterima atau tidak. Penelitian ini menguji bagaimana jenis kelamin dan tipe dilema moral memengaruhi penilaian moral, arousal emosional, dan valensi pada 60 partisipan. Penelitian eksperimen ini menggunakan *factorial design* dan pendekatan *vignette-based scenarios*. Hasil analisis ANOVA faktorial campuran menunjukkan adanya pengaruh utama yang signifikan dari jenis keputusan moral terhadap penilaian, dimana respon deontologis dinilai lebih afirmatif dibandingkan respon utilitarian. Penilaian deontologis menekankan kepatuhan pada aturan atau kewajiban moral tanpa mempertimbangkan konsekuensinya, sedangkan penilaian utilitarian berfokus pada konsekuensi tindakan dan bertujuan memaksimalkan kesejahteraan secara keseluruhan. Jenis kelamin tidak berpengaruh signifikan terhadap keberterimaan moral atau tipe keputusan, namun perempuan menunjukkan tingkat arousal emosional yang lebih tinggi dibandingkan laki-laki. Tipe dilema moral secara signifikan memengaruhi arousal dan keberterimaan moral. Dilema yang melibatkan kerugian tidak disengaja dinilai paling dapat diterima dan menimbulkan arousal paling tinggi. Penilaian valensi menunjukkan skor negatif secara konsisten di semua kondisi, yang mencerminkan sifat konflik moral yang menimbulkan ketegangan emosional. Temuan ini mendukung model *dual-process* dalam kognisi moral, dengan menekankan dominasi faktor emosional dan kontekstual dibandingkan karakteristik individu yang stabil seperti jenis kelamin, dan bermanfaat bagi profesional untuk pengembangan pelatihan etika dan intervensi pengambilan keputusan moral yang kompleks di lingkungan yang sarat emosi atau tekanan emosional.

Kata Kunci: dilema moral, eksperimen vignette, insidental dilema moral, penilaian moral deontologi, penilaian moral utilitarian

Correspondence: Sabiqotul Husna. Email: sabiqotul.husna@uin-suka.ac.id

Morality is a foundational framework for evaluating behavior, which helps in distinguishing between acceptable and unacceptable norms in the society. According to Haidt and Kasebir (2010), it provides a reference for assessing right from wrong. On a daily basis, judgments and behavioral choices are grounded in standards, with morality as a central subject in psychological analyses. Contemporary moral psychology has increasingly examined the ability of individuals to make decisions, particularly through dual-process theory (Greene, 2007). Based on this perspective, the theory posits the existence of two distinct cognitive systems, namely fast, automatic-emotional, and slow, deliberate-rational systems. The first is generally associated with deontological moral judgments, guided by the consistency of an action with related rules or duties, regardless of outcomes. The automatic-emotional system is fast, intuitive, and functions as a result of emotional stimuli. The second system focused on utilitarian judgments, which prioritized outcomes and the maximization of overall well-being. This conscious-controlled procedure, is characterized by the gradual nature, and absolute consideration of the conscious state, as well as a higher cognitive load (Greene et al., 2001; Gawronski & Beer, 2017).

The dual-process theory was used in the description of moral variability across individuals. For example, Conway and Gawronski (2013) reported that deontological judgments were strongly related to empathic concern, religiosity, and perspective-taking, while utilitarian judgments were associated with cognitive deliberation. Similarly, Gleichgerrcht and Young (2013) stated that individuals with higher empathic concern made deontological judgments, and those with lower empathy endorsed utilitarian decisions more frequently. Gamez-Djokic and Molden (2016) stated that deontological responses also evolved from motivational orientations, such as prevention-focused objectives, that were independent of affective cues.

Moral judgments were examined based on aspects of emotional responses that predisposed a person to adopt deontological or utilitarian perspective. Gleichgerrcht and Young (2013) conducted an experimental study and reported that moral judgments correlated with affective aspects, namely empathic concern. Additionally, it was reported that participants who based moral judgments on utilitarian principles had lower empathic concern, compared to individuals who referred to deontological principles. Gamez-Djokic and Molden (2016) stated that individuals passed moral judgments on deontological principles due to intuitive reactions and emotions. This also included motivational orientation, particularly the preventive perception, which failed to engage affective reactions. Conway and Gawronski (2013) also reported that the tendency for moral judgments reliant on deontological principles was related to

empathic concern, religiosity and perspective-taking, while judgments on utilitarian principles was associated with the need for cognition.

The dual-process theory was used in evaluating moral variability by outlining the roles of automatic emotional responses (System 1) and controlled cognitive reasoning (System 2). However, another study reported that these cognitive systems did not operate in isolation from social context. Ellemers et al. (2019) stated that moral behavior cannot be fully understood without considering the situational and motivational forces associated with its emergence. The statement that good people can engage in bad acts outlined the importance of examining the interplay between internal moral reasoning and external pressures. This was in line with the study by Greene (2014) that social framing activated or suppressed specific moral intuitions, influencing the reliance of individuals on deontological (emotion-based) or utilitarian (reason-based) process. Similarly, Cushman (2013) proposed a model in which social feedback and learning histories shaped the weighting of emotional and rational inputs in moral decision-making. The study by (Packer et al., 2021) also reported that social norms and group membership regulated the engagement of dual-processing systems, with individuals exhibiting utilitarian and deontological reasoning when dilemmas included out and in-group targets, respectively. These findings collectively showed that dual-processingsystems were a product of internal mechanisms, as well as dynamically shaped by interpersonal and cultural contexts.

Jin and Peng (2021) conducted a study related to social perception and moral judgment, mainly focusing on the influence of deontological and utilitarian perspectives. The findings showed that moral psychology was influenced by deontological and utilitarian normative ethical theories. Both provided a logical basis for determining morality of behavior. Furthermore, the findings of the four experiments, showed that moral judgments were connected to certain personality types, namely calmness and competency. Deontological and utilitarian moral judgments were associated with warm, and competent personalities, respectively. Another study by McNair et al. (2018) examined the influence of age on deontological and utilitarian moral judgements in respect to decision-making. The participants were divided into two different groups, namely old (55-81 years) and young (18-25 years). The results of the experimental study focused on the differing tendencies to use deontological perspective and idealist beliefs related to moral. Additionally, the older age group had a stronger deontological tendency compared to the younger group.

The integration of social context into moral judgment analyses broadened the knowledge of how individuals navigated complex ethical decisions. This was realized through internal emotional or rational process,

as well as in responses to interpersonal dynamics, group affiliations, and cultural expectations. The socially embedded influences interacted with personal characteristics, such as gender, to shape how moral reasoning was expressed. In particular, gender was the main factor that regulated the influence of emotional and social cues in moral judgment. The intersection between gender and social context resulted in the need for further exploration into how men and women differed in cognitive and affective process when confronted with dilemmas. Therefore, gender differences in emotional processing may partially explain variations in moral decision-making. The meta-analysis of 40 experimental studies by Friesdorf et al. (2015) showed that women engaged in deontological reasoning, while men preferred utilitarian approaches. These differences were connected to stronger affective responses among women, particularly when confronted with moral transgressions associated with harm or danger. The findings were in line with earlier proposed male-centered models, such as the rationalist framework (Kohlberg, 1969). Gilligan (1993) also used the ethic of care model to analyze women's relational, and empathetic orientation towards morality, as opposed to the more abstract and principle-based ethic of justice framework adopted by men. According to Benhabib (1985), men often accepted a detached, third-person perspective, with women engaging in moral reasoning from a first-person, emotionally connected viewpoint.

Following the description above, gender differences in affective processing played a significant role in moral judgment, with women exhibiting stronger emotional reactivity to harm compared to men. The meta-analytic study by Friesdorf et al. (2015) reported that women exhibited significantly stronger deontological inclinations (effect size d = .57), while men showed a modest utilitarian preference (d = .52). The finding suggested that affective responses, rather than purely cognitive evaluations drove these gender-based differences during dilemmas. Similarly, the neuroimaging findings by Harenski et al. (2008) reported that women, exhibited heightened activation in emotion-related brain regions (e.g., the anterior insula and posterior cingulate) when evaluating moral stimuli, thereby exhibiting greater emotional sensitivity. Complementary psychophysiological studies had shown that this gender also experienced stronger startle reflexes and higher perceived unpleasantness in responses to aversive images, confirming biologically rooted sex differences in emotional processing (Bianchin & Angrilli, 2012). These preliminary studies collectively adopted the dual-process model of moral cognition, implying that affective intensity in women substantially shaped moral judgment, with cognitive evaluation performed in the same manner by both genders.

Asides from gender, the engagement type in dilemma significantly influenced moral judgment. The engagement of oneself (self-

involvement) or others (other-involvement) in dilemma often influenced the evaluation of moral decisions. Rybash et al. (1983) investigated how individuals respond differently to dilemmas concerning themselves compared to others. The participants were presented with hypothetical scenarios that required making moral decisions either affecting oneself or other people. The results showed that individuals were generally lenient in accepting morally questionable actions concerning themselves rather than others. This outlined the role of personal engagement in shaping moral judgments, suggesting that self-related dilemmas may be judged with greater empathy or justification compared to decisions affecting others. The study by Sachdeva et al. (2015) explored how self-sacrifice influenced moral decision-making in dilemma scenarios. However, a series of trolley-type dilemmas, were used to monitor whether the subject of the sacrifice was oneself or another person. The experiments conducted on diverse participants, showed that people were willing to endorse utilitarian principles such as making sacrifices to save oneself, rather than another individual. This pattern held across cultural contexts, suggested that moral judgments were sensitive to the outcomes and one's role in dilemma. The study outlined moral value attributed to voluntary self-sacrifice and how personal engagement changed the acceptability of utilitarian decisions.

In this context, the type of dilemma whether incidental or instrumental plays a critical role in shaping moral judgments. Incidental and instrumental dilemmas portray harm as a side effect, and a direct means to an end, respectively. These distinctions influenced both moral acceptability and emotional responses. Lotto et al. (2013) studied how the different dilemmas (incidental vs. instrumental) affected moral decision-making. The results showed that participants rated incidental harms as morally acceptable compared to the instrumental types. Moreover, instrumental dilemmas elicited stronger emotional responses and longer decision times, suggesting greater cognitive and emotional conflict. This outlined the impact of dilemmas on both affective and cognitive aspects of moral reasoning. Ludwig et al. (2020) explored the impact of two main factors, instrumentality (harm is a means to an end) and personal force (harm is caused by direct physical action) on moral judgments in harm-related dilemmas. Process-dissociation approach, was used to assess the participants' deontological and utilitarian inclinations across the different dilemma types. The study found that instrumentality significantly increased deontological responses, suggesting participants were less willing to endorse harm when perceived as a means. Meanwhile, personal force had a weaker and less consistent effect. The findings showed that the structural features of dilemma specifically when harm was used instrumentally played a crucial role in shaping moral reasoning.

Several experimental studies on moral reasoning and judgment had been conducted

in Western countries. Greene et al. (2001) carried out a foundational study using functional neuroimaging during trolley-problem dilemmas. The result showed that emotional brain areas activated during personal dilemmas, was in line with the proposed dual-process model. Hofmann et al. (2014) investigated real-life moral behavior using ecological momentary assessment in a large U.S. sample (N = 1,252). The study found that moral and immoral experiences were frequently shaped by contextual factors, such as political orientation. Furthermore, moral contagion and licensing, outlined that decision-making was influenced by internal reasoning and dynamic social contexts. A recent large-scale cross-cultural study by Doğruyol et al. (2019), using 5,317 participants from 20 countries, including those categorized as WEIRD societies such as the United States, Canada, the United Kingdom, and Australia, as well as non-WEIRD nations namely Turkey, Brazil, India, and Kenya aslo showed interesting findings. It was reported that the five core moral dimensions Care, Fairness, Loyalty, Authority, and Sanctity were stable across both WEIRD and non-WEIRD cultural contexts. The term *WEIRD* referred to Western, Educated, Industrialized, Rich, and Democratic societies, which had been disproportionately represented in psychological analyses. However, *non-WEIRD* societies focused on collective values, comprehensive reasoning, and relational morality. The study outlined the cross-cultural robustness of moral reasoning patterns and importance of including non-WEIRD populations, by showing the structural consistency of moral foundations across the diverse contexts.

In some Asian countries, prior studies on moral judgment had made significant progress, despite the need for broader and deeper investigation across the region. Zhang and Thoma (2016) conducted a cross-cultural study on moral judgment development in mainland China, engaging 300 participants across diverse age groups. Mixed-methods including the combination of quantitative moral dilemma questionnaires based on Kohlberg's Moral Judgment Interview with qualitative dialogues were adopted to deepen the understanding of participants' reasoning. The findings showed that Chinese participants adhered to a similar moral development progression as Western populations but exhibited Confucian values such as social harmony, respect for authority, and collectivism. The participants prioritized interpersonal relationships and social obligations, compared to the Western samples. This suggested that the moral development structure was universal, and the content shaped by cultural context. Qian et al. (2024) examined moral judgment differences between Japanese and Chinese participants, with a focus on cultural and gender influences using the CNI (Consequences, Norms, and Inaction) model. The study included a balanced sample of male and female young adults from both countries

and used dilemma related tasks associated with separately measuring sensitive consequences, adherence to moral norms, and preference for inaction. The findings proved that Japanese participants had higher sensitivity to moral norms and a stronger preference for inaction compared to the Chinese, who were focused on the consequences of the actions. Meanwhile, gender differences were also observed, with women from both countries prioritizing moral norms. This study centered on how cultural values and gender differences collectively shape moral decision-making process. The finding showed that moral judgment was influenced by individual cognitive mechanisms and culturally embedded social expectations.

Some of the previous studies conducted in Asian countries, led to considerable scope for further investigation. This was particularly within the Indonesian context, where reviews on moral judgment were relatively limited. Aurelia et al. (2025) conducted a cross-cultural study comparing moral judgments of corruption between participants from the Netherlands and Indonesia. Process comprised a total of 600 participants, evenly divided between the two countries. Experimental vignettes were used to describe various corruption scenarios to assess moral condemnation and acceptability. Additionally, the participants rated the severity and moral injustice of each scenario, enabling the examination of cultural differences in respect to corruption. The results showed that the Indonesian participants tended to judge corruption more harshly than the Dutch, reflecting stronger moral condemnation in this context. Cultural values, such as collectivism and individualism practiced in Indonesia and the Netherlands, respectively influenced these moral evaluations. The study further outlined how cultural context shaped moral judgment, particularly complex social issues such as corruption. Oktrivina et al. (2024) investigated ethical decision-making among professional accountants, focusing on how moral judgment mediated the relationship between ethical awareness and actual decision outcomes. Based on this perspective, a quantitative approach, was used to carry out a survey on 150 professional accountants in various organizations. The findings showed that moral judgment played a critical mediating role, influencing how accountants translated ethical awareness into practical decisions. Furthermore, it outlined the significance of strengthening moral reasoning skills to improve ethical conduct within the accounting profession. Another study by Ji et al. (2009) examined the relationship between Islamic personal religiosity and moral reasoning related to social justice and equality among Indonesian college students. The study consisted of 200 participants selected from diverse universities, using a quantitative survey design. The levels of personal religious commitment and moral reasoning were assessed through standardized questionnaires. The findings proved that

stronger Islamic religiosity positively correlated with higher levels of moral reasoning concerning social justice and equality issues. Moreover, it outlined the significant influence of personal religious beliefs on moral development within the Indonesian cultural context.

The limited number of studies conducted in Indonesia, resulted in the need for further investigations within this context. Therefore, the present study aimed to explore whether gender differences (male and female) influenced adults' responses to dilemma scenarios, particularly in relation to deontological and utilitarian principles. The potential differences in relation to affective reactions between male and female adults during moral decision-making, were also examined. This led to the following proposed five hypotheses based on dual-process theory. H1: Gender influenced affirmative responses to dilemmas, with women and men likely to make deontological and utilitarian decisions, respectively. H2: Gender affected moral acceptability ratings, where women judged deontological decisions as less acceptable, while men perceived utilitarian decisions as more acceptable. H3: Gender differences were observed in affective responses (arousal and valence) during moral judgment, with women showing stronger emotional reactions than men. H4: Moral judgments differed between self and other-involvement dilemmas, with self-participation dilemmas regarded as more acceptable. H5: Moral judgments differed between instrumental and incidental dilemmas, with incidental type considered more acceptable.

Method

This study adopted a factorial experimental design using vignette-based scenarios to explore moral judgment. The factorial design allowed the examination of multiple independent variables (IVs) and its interactions with several dependent variables (DVs). A 2 (Gender: male vs female) × 2 (Type of moral dilemma: instrumental vs incidental) × 2 (Involvement: self-involvement vs other-involvement) was used between-subjects factorial design. Among the three IVs, dilemma types and participation were experimentally manipulated, while gender was a measured variable, resulting in the overall mixed factorial design (a measured variable was observed, recorded, or categorized as it exists). Participants were randomly assigned to the order of presented dilemma conditions. In this context, each participant was presented with all 26 dilemma scenarios derived from the 2×2×2 factorial design. This method allowed the assessment of the main effects, as well as interactions among cognitive, emotional, and contextual factors influencing moral decision-making.

The details of the variables were stated as follows 1) IV (IV/factor) Gender (female vs male), 2) IV type of dilemma (instrumental vs deontological theme), 3) IV type of dilemma

(self-involvement Vs involvement of others). This led to the inference that there were 2 (gender: male vs female) X 2 (moral dilemma: instrumental vs incidental) X 2 (self-involvement vs involvement of other people) in the context of the IV. The DV consisted of 1) DV 1: affirmative responses to dilemma (decision to act or not), 2) DV 2: moral acceptability rating (assessment of moral acceptability; whether it is acceptable or not), 3) DV 3a: arousal rating (a measure of arousal; calm or anxious in deciding the affirmative responses), 4) DV 3b: valence rating (a measure of the valence of feelings: positive/pleasant or displeased/unpleasant when deciding on an affirmative responses).

Table 1

Study Design

Factors	Measurements
Gender (male vs female) Moral dilemma (instrumental vs incidental) Moral dilemma (self-involvement vs other-involvement)	Affirmative response Moral acceptability rating Arousal rating Valence rating

The participants were mostly adults, both males and females, aged between 18 and 40 years. Additional inclusion criteria were have never participated in dilemma experiment, possessed normal vision or corrected-to-normal with glasses, and no difficulty reading text on a computer screen. In line with the study, sampling was conducted using a quota technique. The sociodemographic control factors consisted of education level (undergraduate/graduated, yes/no), and ethnicity (Javanese, yes/no). Participants were recruited voluntarily by disseminating information related to the experiment through social media and broadcast messages. These individuals were divided into two groups, namely male and female. A power analysis using G*Power 3.1 showed that a minimum of 34 participants were required to detect medium effect sizes (f = .25), with 80% power in repeated-measures factorial ANOVA. Meanwhile, the participants comprised mainly of students from the universities in Yogaykarta and Java. This limited the generalizability of findings, and sampling resulting in the need for the use of diverse populations in future studies. To minimize potential order effects, counter balancing was implemented by randomizing the presentation sequence of dilemma scenarios across participants. This was achieved using the randomization tool available at https://randomizer.org (Urbaniak & Plous, 2013).

The two groups (male and female groups) of participants, were treated in a similar manner. The order in which the diverse scenarios were shown was randomized for each participant, to control order effects and reduce potential biases caused by the sequence

of presentation. The randomization process was performed using the randomizer.org site (Urbaniak & Plous, 2013) to generated the eight unique, random sequence scenarios. However, every participant witnessed the same set of scenarios in a different, randomly determined order. This method ensured that responses were not influenced by the position of a scenario within the sequence, improving the internal validity of the experiment. After debriefing and signing of informed consent, the participants individually received a trial of answering moral dilemmas (three scenarios) accompanied by an assistant. Furthermore, a test was conducted, where 26 scenarios (Lotto et al., 2013) were presented on a computer screen in a text format.

The scenarios were administered using the Labvanced platform (Labvanced, 2025), an online experimental software that allowed for precise control over stimulus presentation and responses collection. Each participant completed the task individually, seated in front of a standard-sized computer screen in a quiet laboratory setting. The scenarios were carefully read and then decisions made by clicking on the responses option at the end. These were showed in randomized order to control sequence effects, with each participant exposed to all 26 scenarios. On average, the task required approximately 30 to 45 minutes to complete, depending on reading speed and responses time. The participants proceed at own pace but instructed to avoid revisiting previous questions thereby ensuring natural and intuitive responses. The digital format enabled automated data collection, enhancing both accuracy and efficiency in recording the diverse moral judgments. All participants engaged in the experiment in a closed laboratory guided by the assistants. This study was evaluated and approved by the Institutional Review Board with approval ethical clearance number B-parti1583.1/ Un.02/L3/TU.03/06/2023. Additionally, the participants willingly agreed to participate in the experiment, after being informed about all risks, including the right to withdraw, future publication of the data and procedures required for the signing of the consent form.

The instruments used were adopted from established studies, although these were not subjected to formal validation in the Indonesian cultural context. Cognitive interviewing was not conducted during development process, and the omission was perceived as a methodological limitation. This was because cognitive interviews played a crucial role in ensuring that participants interpreted items consistently and appropriately across cultural contexts. Future studies should incorporate a full adaptation process, including translation–back and pretesting. Dilemma presented described moral issue related to the choice to save the lives of more people, by sacrificing one or two lives. The participants were asked to decide on moral judgment for these scenarios (Table

1) by responding to two questions: 1) Would you take the action proposed in the scenario? (Affirmative responses; Yes or No), 2) How morally acceptable do you think the action taken was? (Moral acceptability; followed by a scale from 1 (not at all acceptable) to 7 (completely acceptable). High scores represented utilitarian moral judgment rather than deontological. The participants were also asked to rate own affective valence (from a scale of 1 to 7) and arousal (from a scale of 1 to 9) when deciding on moral judgments. This included rating how calm or anxious and unpleasant or pleasant the participants felt when making moral decisions (Self-Assessment Manikin/SAM; Lang et al., 2008).

Table 2

Sample Instrumental, Incidental, and Filler Dilemmas (Lotto et al., 2013)

Dilemma	Scenario	Resolution
Incidental Other-involvement	Imagine being a fighter jet pilot on a mission in Southeast Asia. During the flight, you notice that another fighter jet on the same mission had accidentally launched a missile at a light aircraft carrying six people. You notice that there was also another military reconnaissance aircraft in the same area.	A procedure was initiated to destroy the missile, based on the awareness that when it exploded, the military recon naissance aircraft would crash, leading to the death of the pilot. However, the six people on board the light aircraft would definitely sur vive.
Incidental Self-involvement	Supposing you are traveling with two of your friends, one of whom is driving, and the car is on the right lane. Suddenly, the driver becomes unconscious and the car starts to accelerate automatically. The driver of the truck uses the brakes, despite the unavoidable impact. To the right, is a lay-by where a road worker is working.	You immediately take over the steering and swerve into the lay-by to avoid the truck in front. Meanwhile, you are aware that the worker in the lay-by will be hit and killed, but you and your two friends would survive.
Instrumental Other-involvement	You are in charge of a special team of technicians performing maintenance work on a large elevator. Six technicians are working in the elevator shaft. Additionally, you and another technician are on the top floor in the engine room. The hoist breaks and the elevator cabin starts falling.	You push the technician next to you into the mechanism wheels so that his body causes the elevator to stop. Moreover, you are aware that he would die, but the other six technicians are bound to survive.
Instrumental Self-involvement	Supposing you are the fourth person on a team of five mountain climbers engaged in a climb. The team leader had just secured himself to the rock face when the second person starts to slide, pulling you, and the others, with him. You all fall tens of meters and come to a stop above a cliff. Besides, you are too heavy and the rope will not hold for long.	To lighten the load, you decide to cut the rope connecting you to the last climber. You are completely aware that he will fall into the cliff and die, but the other climbers would survive.

Data obtained from the experiment were analyzed using Repeated Factorial Analysis of Variance (ANOVA). This was aimed to determine whether there were mean differences between groups (men x women) regarding deontological and utilitarian moral judgments (affirmative responses and moral acceptability). Furthermore, process included affective reactions (arousal and valence) and mean differences between the diverse scenario types (instrumental x incidental).

Results

In this study, the 60 participants were divided into female and male groups, with both containing 30 individuals. Table 3 shows the distribution and percentage of the participants.

Table 3

Participant Demographic Data

Characteristic	%
Age	
< 20 years old	20
21- 26 years old	73.3
> 27 years old	6.7
Gender	
Male	50
Female	50
Current Education	
Undergraduate student	88.3
Graduate student	3.4
Graduate from bachelor's degree	6.7
Not student	1.6
The Origin of Education Institution	
State Islamic University	75.6
Sunan Kalijaga	8.3
Gadjah Mada University	7.7
Nahdlatul Ulama University	6.7
State University of Yogyakarta	1.7
Surya Global School of Health Science	
Race	
Javanese	80
Sunda	8.4
Javanese – Sunda	6.5
Javanese – Aceh	1.7
Padang	1.7
Dayak Paser	1.7

Note. N = 60. M = 22.45.

In the context of age, the participants were predominantly aged 21-26 years, constituting 73.3% equivalent to 44 individuals. However, 20% or 12 participants were aged 20 years and below. The minority participants in the experiment consisted of 6.7% or 4 people aged 27 years and above. In terms of gender, 50% of men, and women, was equivalent to 30 participants. As for the education context, a total of 88.3% or 53 participants (the majority) were active students at the undergraduate level (S1). Additionally, 6.7% or fourindividuals were active master students (S2).

Assumption tests for normality and homogeneity of variance were conducted, prior to hypothesis testing. Moreover, due to violations in normality in some cells, bootstrapped 2×2×2 mixed factorial ANOVA was applied where appropriate. During the main analysis, the effect of gender (between-subjects), and moral decision type (within-subjects: deontological vs. utilitarian) on affirmative responses to dilemmas (hypothesis 1), was examined.

Table 4

Descriptive Statistics of Moral Judgment by Gender

Gender	M Deontological	M Utilitarian	M Difference (Deontological - Utilitarian)	SD	N
Female	17.8	13.4	4.4	3.1	30
Male	16.2	14.7	2.5	2.8	30

Note. Mean Difference represented the average score between deontological and utilitarian moral decisions.

Table 4 shows the descriptive statistics of mean differences in moral decision types based on gender. The mean difference score implied that female participants (mean difference = 4.4, SD = 3.1) were more inclined to make deontological decisions than males (mean difference= 2.5, SD = 2.8). This was tested statistically by conducting a 2 (Gender: Male vs. Female) × 2 (Decision Type: Deontological vs. Utilitarian) mixed factorial ANOVA on affirmative responses.

Table 5

Repeated Measures Factorial ANOVA (Affirmative Responses)

Effect	SS	Df	MS	F	*p*	η^2_p
Decision Type	186.30	1	186.30	28.45	< .001	.33
Decision × Gender	24.70	1	24.70	3.77	.057	.06
Residual	379.90	58	6.55	—	—	—

*lote***.** *N* = 60. Decision Type = deontological vs. utilitarian judgment. SS = Sum of Squares; MS = Mean Square. *p* values are eported to three decimal places unless < .001, η^2_p = partial eta squared.

Table 5 shows that the main effect of moral decision type on participants' scores, $F(1, 58) = 28.45$, $p < .001$, $\eta^2_p = .33$. According to the study by Cohen (1988), the result represented a large effect size. This suggested that majority of the participants, provided affirmative responses to deontological dilemmas, regardless of gender. The interaction between decision type and gender was not statistically significant, $F(1, 58) = 3.77$, $p = .057$, $\eta^2_p = .06$, reflecting a small-to-medium effect size. The interaction did not reach conventional significance levels. The marginal p-value represented a potential trend that female participants engaged in deontological reasoning slightly more than the male counterparts, although this difference was not definitive. The findings supported the inference that moral decisions were strongly influenced by the type of moral reasoning (deontological vs. utilitarianism) rather than gender. The inclusion of effect sizes showed the magnitude beyond statistical significance.

Knowledge of how individuals judged moral acceptability of various dilemma types was crucial in comprehending the inherent psychological and social mechanisms that shaped ethical decision-making. The different dilemmas types, namely incidental harm, instrumental actions, self-involvement, or the involvement of others, triggered distinct patterns of moral evaluation. Additionally, gender differences influenced reasoning styles, potentially leading to different sensitivity levels to moral conflict. Regarding these possibilities, dilemma types and gender of the participants reportedly affected moral acceptability judgments. To establish the empirical basis for further statistical testing, the following descriptive analysis summarized the mean and standard deviation of moral acceptability ratings across four dilemma types and two gender groups.

Table 6

Descriptive Statistics of Moral Dilemma Type and Gender (Moral Acceptability)

Moral Dilemma Types	Female		Male		Total	
	M	*SD*	*M*	*SD*	*M*	*SD*
Incidental	46.40	16.53	43.87	15.25	45.13	15.88
Self-involvement	40.93	16.54	38.93	15.93	39.93	16.19
Others' Involvement	39.53	15.91	36.87	15.40	38.20	15.65
Instrumental	35.13	17.45	32.90	17.67	34.02	17.51

Descriptive analysis showed that *incidental* dilemmas had the highest moral acceptability ratings ($M = 45.13$, $SD = 15.88$), followed by *self-involvement* ($M = 39.93$, $SD = 16.19$), *others' Involvement* ($M = 38.20$, $SD = 15.65$), and *instrumental* dilemmas ($M = 34.02$, $SD = 17.51$). In addition, these patterns were largely consistent across genders.

The repeated measures factorial ANOVA was conducted to examine whether participants' gender and dilemma type influenced moral acceptability ratings

(Hypotheses 2, 4 & 5). The within and between-subject factors were *Moral Dilemma Types* (Instrumental, Incidental, Others' Involvement, and Self-involvement), and *Gender* (Male vs. Female), respectively.

Table 7

Repeated Measures Factorial ANOVA on Moral Acceptability Ratings by Moral Dilemma Type and Gender

Source	SS	Df	MS	F	*p*	η^2_p
Moral Dilemma Type	36.524.87	3	12.174.96	140.12	<.001	.71
Gender	218.45	1	218.45	1.23	.272	.02
Moral Dilemma Type × Gender Interaction	384.36	3	128.12	1.47	.227	.03

Note. SS = Sum of Squares. *MS* = Mean Square. *F* = F-rasio. *p* = significance value. η^2_p = partial eta squared. The within-subject factor was Moral Dilemma Type (Instrumental, Incidental, Others' Involvement, Self-involvement), and the between-subject factor was Gender (Male, Female).

The results of the ANOVA in table 7, showed the significant effect of Moral Dilemma Type, $F(3, 174) = 140.12$, $p < .001$, $\eta^2_p = .71$. This implied that dilemma type significantly affected moral acceptability. However, Gender, $F(1, 58) = 1.23$, $p = .272$, $\eta^2_p = .02$, had no significant effect. In this context, no significant interaction existed between dilemma type and gender, $F(3, 174) = 1.47$, $p = .227$, $\eta^2_p = .03$. The findings suggested that gender did not influence how participants judged the acceptability of the actions in each dilemma type.

Post-hoc Bonferroni-corrected comparisons were performed, to completely understand the differences between dilemma types. The results in table 8. showed that *Incidental* dilemmas were rated as significantly more acceptable than all other types (all $p < .001$), while *Instrumental* dilemmas received the lowest acceptability ratings. *Others' Involvement* and *Self-involvement* dilemmas were rated moderately and did not significantly differ from each other ($p = .298$). These findings proved that participants abruptly distinguished between unintentional (Incidental) and intentional or instrumental harm, regardless of whether it concerned themselves or others.

Table 8

Post-Hoc Pairwise Comparisons (Bonferroni-Corrected)

Comparison	M Difference	*p*
Incidental vs. Instrumental	+11.11	< .001*
Others' Involvement vs. Instrumental	+4.18	.004*
Self-involvement vs. Instrumental	+5.91	< .001*
Incidental vs. Others' Involvement	+6.93	< .001*
Incidental vs. Self-involvement	+5.20	.001*
Others' Involvement vs. Self-involvement	-1.73	.298

Note. *p = < .001

Moral acceptability was strongly influenced by dilemma type, rather than gender. Participants were most forgiving of Incidental harm and disapproving of harm used instrumentally, reflecting nuanced moral evaluations across different contexts. Furthermore, a factorial ANOVA was conducted to examine the third hypothesis, which proposed gender had an effect on the level of arousal experienced when making decisions regarding dilemmas. This analysis aimed to determine whether there were statistically significant differences in arousal scores between male and female participants across various dilemma types.

Table 9

Factorial ANOVA on Arousal by Moral Dilemma Type and Gender

Effect	SS	Df	MS	F	*p*	η^2_p
Moral Dilemma Type	5600	3	1866.7	3.45	.017	.043
Gender	10062	1	10062.0	18.65	<.001	.074
Moral Dilemma Type × Gender Interaction	147	3	49.0	.91	.438	.012

Note. SS = Sum of Squares. *MS* = Mean Square. *F* = F-rasio. *p* = significance value. η^2_p = partial eta squared. The within-subject factor was Moral Dilemma Type (Instrumental, Incidental, Others' Involvement, Self-involvement), and the between-subject factor was Gender (Male, Female).

The result in table 9. showed the significant effect of gender, $F(1, 232) = 18.65$, $p < .001$, $\eta^2_p = .074$, with female participants reporting higher arousal scores than the male counterparts. Furthermore, dilemma type, $F(3, 232) = 3.45$, $p = .017$, $\eta^2_p = .043$, had a significant effect implying that arousal scores differed across the four types. The interaction between gender and dilemma type was not significant, $F(3, 232) = .91$, $p = .438$, suggesting that the effect was consistent across all dilemma types.

Considering that the main effect of dilemma type on arousal was statistically significant in the ANOVA analysis, a follow-up post hoc test was conducted to identify specific types differed significantly in eliciting emotional arousal. The aim of the post hoc test was to examine pairwise differences between the four dilemma types, namely incidental, instrumental, self-involvement, and others' involvement in order to determine which scenarios generated stronger emotional responses.

Table 10

Post Hoc Comparisons – Types of Moral Dilemmas

Comparison	*M*	SE	*p*	95% CI	
				LL	UL
Incidental vs Instrumental	.82	.23	.002*	.29	1.35
Incidental vs Others' Involvement	.95	.25	.001*	.38	1.52
Incidental vs Self-involvement	.38	.22	.092	-.04	.80
Self-involvement vs Instrumental	.44	.21	.038*	.02	.86
Self-involvement vs Others' Involvement	.57	.23	.014*	.12	1.02
Instrumental vs Others' Involvement	.13	.20	.510	-.26	.52

Note. CI = Confidence Interval. LL = Lower Limit. UL = Upper Limit. *$p < .05$

The post hoc comparisons in table 10. showed that incidental dilemmas prompted significantly higher arousal than both instrumental (p = .002) and others' involvement (p = .001), implying the scenarios provoked stronger emotional responses. Although incidental dilemmas also produced greater arousal than self-involvement, this difference was not statistically significant (p = .092), showing comparable levels of affective engagement. Self-involvement dilemmas prompted significantly more arousal than instrumental (p = .038) and others' involvement dilemmas (p = .014), further supporting the idea that moral situations perceived as personally relevant or concerning oneself tended to activate stronger emotional arousal. However, there was no significant difference in arousal between instrumental and others' involvement conditions (p = .510), showing that these two dilemma types may be less emotionally activating compared to the more personally engaging scenarios.

Factorial anova was also conducted to test the effects of the interaction between gender, and dilemma type on valence.

Table 11

Factorial ANOVA on Valence by Moral Dilemma Type and Gender

Effect	SS	Df	MS	F	*p*	η^2_p
Moral Dilemma Type	288	3	96.1	.192	.902	.002
Gender	917	1	916.5	1.827	.178	.008
Moral Dilemma Type x Gender	237	3	78.9	.157	.925	.002

Note. SS = Sum of Squares. *MS* = Mean Square. *F* = F-rasio. *p* = significance value. η^2_p = partial eta squared. The within-subject factor was Moral Dilemma Type (Instrumental, Incidental, Others' Involvement, Self-involvement), and the between-subject factor was Gender (Male, Female).

Based on the analysis, table 11. showed the variables had no statistically significant effects. The main effect of gender was non-significant, $F(1, 232) = 1.83$, $p = .178$, $\eta^2_p = .008$, showing that male and female participants did not differ meaningfully in respective emotional valence ratings. Similarly, the main effect of dilemma type was non-significant, $F(3.232) = .19$, $p = .902$, $\eta^2_p = .002$, suggesting that the specific type whether incidental, instrumental, self-involvement, or others' involvement—did not significantly influence valence. The interaction between gender and dilemma type was also not significant, $F(3.232) = .16$, $p = .925$, $\eta^2_p = .002$. Although the effects were statistically non-significant, the descriptive data showed that the mean valence scores across all conditions ranged between 3.1 and 3.5 on a 7-point scale (where 1 = very unpleasant, 7 = very pleasant, and 4 = neutral). This implied that, overall, participants rated dilemmas as rather unpleasant, suggesting the scenarios prompted mild negative emotional responses regardless of gender or dilemma type. The tendency was consistent with the emotionally conflicting nature of dilemmas, which often included ethical ambiguity, personal risk, or harm and factors commonly associated with negative affect.

Discussion

This study explored how gender and different dilemma types influenced moral judgment (affirmative responses and moral acceptability), emotional arousal, and valence. Several findings arising from the data, focused on the psychological dynamics associated with moral decision-making. The first hypothesis (H1) posited that gender influenced affirmative responses to dilemmas, based on dual process theory. This caused the females to be more inclined to deontological moral decisions, while males tended toward utilitarian reasoning. The results partially supported this hypothesis, because moral decision type had a significant effect, with participants generally exhibiting a stronger preference for deontological over utilitarian decisions ($\eta^2_p = .33$). Additionally, the finding showed that, regardless of gender, individuals tended to reject actions that violate moral rules, even when it could result in greater benefits. This was in line with the dual process theory, which suggested deontological judgments evolved from intuitive, affect-driven process, and utilitarian decisions required more effortful, controlled cognitive reasoning (Greene et al., 2001; Haidt, 2001). Although the interaction between gender and decision type was not statistically significant ($p = .057$), the trend approached significance and showed a small-to-moderate effect size ($\eta^2_p = .06$). The female participants preferred deontological judgments slightly more than the males. This supported prior findings that women were more emotionally responsive and averse to causing harm, leading to bias toward non-utilitarian outcomes (Fumagalli et al., 2010). The trend did not meet the conventional thresholds for statistical

significance, rather it offered preliminary evidence for gender-based differences in moral cognition, leading to the further exploration of more diverse samples.

Building upon the description above, gender-based trends should be interpreted with caution. Hyde's Gender Similarities Hypothesis (Hyde, 2005) stated that gender differences associated with psychological traits, including moral reasoning, were negligible, and often fluctuated across the lifespan of an individual. Hyde further stated that several observations were confounded by contextual variables such as age, culture, and situational factors. When these variables were controlled, men and women often exhibited similar patterns of moral reasoning. The finding was in line with the study by Graham et al. (2009) that although traditional views associate women and men with care and justice-based reasoning, respectively both genders relied on a shared set of moral intuitions. Based on this perspective, gender differences in moral judgment were highly context-dependent. Hypothesis 1 was partially supported, because both gender exhibited deontological reasoning. The females showed a modest, although not statistically significant tendency to strongly endorse related responses. The pattern reflected broader theoretical and empirical debates, outlining the potential for subtle gender influences on moral judgment. This also entailed the overarching similarities between men and women in moral cognition.

Evolutionary Psychology suggested that several behaviors, including moral decision-making, were shaped by pressure. Preliminary studies reported that men and women adopted different moral decision-making strategies (e.g., women focusing on care, and men on justice), and this view perceived as overly simplistic. Certain studies reported that gender differences in moral decisions arose under specific evolutionary pressures, not biological inherent. When environmental and societal conditions change, evolutionary pressures on moral behavior converges, reducing or eliminating gender-based differences (Buss, 2019). Gendered expectations associated with traits such as compassion (associated with women) or assertiveness (associated with men) may not lead to different moral decisions, specifically when individuals step outside these roles. In environments where both genders were prompted to express empathy and justice equally, moral decision-making was less influenced by this variable (Eagly & Wood, 2012). Gendered behaviors, including moral reasoning, were exhibited through socialization and cultural expectations. Therefore, when these external forces are neutralized (e.g., in gender-equitable environments), similar moral decisions are taken by both men and women. The difference in moral reasoning was often attributed to social roles rather than innate dispositions (Ridgeway, 2001).

Hypothesis 2 (H2) stated that gender influenced participants' acceptability ratings of moral decisions, and this was in line with dual

process theory. It was hypothesized that female participants were inclined to deontological reasoning, and rejected harmful actions regardless of outcomes, resulting in lower acceptability ratings. However, the male counterparts were more utilitarian, showing greater willingness to accept harmful actions, leading to higher acceptability ratings. The results did not support this hypothesis, because the factorial ANOVA showed gender, $F(1.232) = 1.12$, $p = .29$, $\eta^2_p = .005$, had no significant effect and interaction with moral dilemma type, $F(3.232) = .67$, $p = .57$, $\eta^2_p = .009$. It suggested that both male and female participants judged the acceptability of dilemmas in statistically similar ways, regardless of the specific moral scenario. Based on the existence of non-significant trend, particularly in emotionally triggered dilemmas, the absence of meaningful gender effects showed that moral acceptability judgments were universally driven by features of the scenario. This included intentionality and perceived harm, rather than by demographic variables such as gender. The outcome challenged some assumptions of dual process theory regarding gender-based moral intuitions, resulting in a more nuanced exploration of individual and contextual factors that shape moral reasoning.

Moral Foundations Theory (MFT) proposed by Haidt and Kasebir (2010) stated that moral reasoning was based on several innate psychological foundations, such as harm, fairness, loyalty, authority, and sanctity. MFT also acknowledged certain gender-based differences (e.g., women might focus more on care and harm, men on justice and fairness), and also outlined the universal nature of these moral foundations. Moreover, the theory suggested that gender differences in moral acceptability may arise from societal and cultural diversities rather than inherent psychological changes. Both men and women shared the same moral basis, weighed differently with no fundamental division in the acceptability of such behaviors. Over time and in different contexts, these differences tended to diminish, specifically when individuals had similar exposure to societal norms (Graham et al., 2009). In diverse cultures, the gap in moral reasoning may be negligible, or not exist at all, implying that the perceived differences between men and women in decision-making was not universal. However, in societies where both genders were equally exposed to similar moral education and role models, differences in decisions tended to diminish (Wei, 2023).

Hypothesis 3 (H3) stated that gender differences were observed in affective reactions, specifically arousal and valence, during moral decision-making, and the female participants experienced stronger emotional reactions. In addition, the hypothesis was grounded in dual process theory, which focused on the role of emotional processing particularly in deontological responses. This was also in line with prior findings that females tended to show greater emotional reactivity to moral and social

stimuli (Domes et al., 2010; Fumagalli et al., 2010). The results partially supported H3, particularly with respect to arousal, where female participants experienced significantly higher arousal levels than the male counterparts, $F(1.232) = 18.49$, $p < .001$, $\eta^2_p = .074$, showing a moderate effect size. The finding suggested that women experienced greater emotional intensity when processing morally complex or ethically charged scenarios. The heightened arousal also outlined the observed (though non-significant) tendency towards making more deontological judgments, consistent with the affective intuitionist model proposed in dual process theory (Greene et al., 2001; Haidt, 2001). Analyzing this difference, from a biopsychological view showed that due to the varying mechanisms of hormonal fluctuations between both gender, women experienced a greater level of arousal. According to McLean and Anderson (2009) differences in emotions and arousal between men and women was due to variations in the mechanisms of hormone fluctuations (women's hormones fluctuate in every 28-day cycle, while in men it remained constant). Women tended to possess higher levels of anxiety due to differing composition of neurotransmitters in the brain. Harensi et al. (2008) carried out a study on moral sensitivity, and further reported that both gender had differing neural mechanisms, leading to varying levels of emotion during moral decisions. Dilemma type significantly influenced arousal levels, $F(3.232) = 3.49$, $p = .017$, $\eta^2_p = .043$. Additionally, incidental harm dilemmas triggered the highest arousal, due to its unexpected and ambiguous nature. In this case, harm occurred without direct intention, inducing strong emotional dissonance. Self-involvement dilemmas also generated increased arousal, suggesting that scenarios with personal stakes provoked intense emotional responses compared to more detached, instrumental or others' involvement situations. Post hoc tests showed that incidental dilemmas produced significantly greater arousal than instrumental ($p = .032$) and others' involvement ($p = .041$), but slightly differed from self-involvement ($p = .087$). These findings proved that both personal relevance and perceived ambiguity were the main drivers of emotional activation in moral contexts.

Hauser et al. (2007) examined how emotional arousal affected moral judgment, proving that it was the main factor in moral decision-making. The study focused broadly on the role of emotions, and also suggested that gender differences contributed to variations in arousal during moral dilemmas. Higher levels of emotional arousal were connected to more deontological (rule-based) judgments, while lower levels were associated with more utilitarian (consequentialist) decisions. Gender differences centered on how emotional arousal influenced moral reasoning, with women experiencing higher levels, and making moral judgments grounded in deontological ethics. The men experienced less emotional arousal,

tending towards utilitarian decisions, despite exhibiting deontological reasoning. Bechara (2004) stated that gender differences in emotional arousal influenced the ways individuals approach moral judgments. Women were more sensitivity to emotional cues, resulting in heightened arousal and moral judgments, particularly in cases concerning personal harm or social justice issues. The men were less influenced by emotional arousal. Cushman et al. (2006) studied the role of affective responses (emotional reactions) in moral decision-making and reported that women showed stronger emotional reactions to dilemmas associated with personal harm or social injustice, leading to stronger deontological (rule-based) judgments.

The valence data did not support H3, because gender or dilemma type had no significant effects on valence ratings (*p*s > .05). However, descriptive statistics showed a consistent pattern of slightly unpleasant emotional reactions, with mean valence scores ranging from 3.1 to 3.5 on a 7-point scale (1 = very unpleasant, 7 = very pleasant). This suggested that participants, regardless of gender or dilemma type, experienced a negative affective tone during dilemmas. The emotional discomfort was expected, considering the ethically problematic nature of the scenarios presented, including potential harm, personal conflict, and moral transgression. The valence responses remained stable across groups, with the arousal data supporting the notion that women were more emotionally reactive during moral evaluations. This partially validated Hypothesis 3, reinforcing the role of emotion in moral cognition. According to Carmona-Perera et al. (2013), individuals who perceived pleasant and unpleasant images as more pleasant and less unpleasant, respectively tended to make more utilitarian moral decisions. This suggested that lower emotional sensitivity to negative stimuli (reduced unpleasantness) facilitated cost–benefit reasoning in dilemmas. Based on the current findings, valence ratings remained consistently negative, with most participants maintaining strong affective responses to morally troubling scenarios, thereby inhibiting utilitarian choices. This was in line with the idea that valence reflected general emotional engagement with moral conflict, and variations in sensitivity partly explained individual differences in reasoning styles. Additionally, McDonald et al. (2021) conducted a meta-analysis examining how valence framing the emotional tone with which moral information was presented affected judgment. The study reported that positively or negatively framed scenarios significantly shifted moral evaluations, even when the underlying content remained constant. This showed the sensitivity of moral judgment to emotional context. In connection to the present study, where valence ratings were consistently negative across gender and dilemma types, the finding suggested that the inherent negative framing of dilemmas

(involving harm or conflict) contributed to the stable pattern of unpleasant emotional reactions. The findings reinforced the idea that valence in moral decision-making was shaped by how scenarios were emotionally framed compared to stable individual traits, such as gender.

The findings were in line with Hypothesis 5 (H5), which stated incidental dilemmas were rated as being more acceptable than the instrumental type. The results showed dilemma type had a substantial effect on moral acceptability ratings ($\eta^2_p = .71$), with incidental harm scenarios judged as significantly more acceptable compared to the instrumental type. This pattern showed the importance of intentionality and causality in moral evaluation. For example when harm occurs as an unintended side effect, participants were more lenient in respective moral judgment compared to when used intentionally as a means to achieve a goal. The findings were consistent with the principle of double effect and previous studies which reported people drew sharp moral distinctions between foreseen and intended consequences (Cushman et al., 2010). Therefore, H5 was strongly supported, and the result was in line with study by Lotto et al. (2013) that incidental dilemmas were more accepted than the instrumental type. When asked to sacrifice the lives of one or two persons to save more people during unplanned situations, the participants agreed. Greene et al. (2004) studied the psychological and neurological basis of moral decision-making, differentiating between instrumental (actions intended to cause harm as a means to an end) and incidental harm (harm that occurs as a byproduct of an intended action with no detrimental intent). The findings suggested that incidental harm was judged as less morally reprehensible than the instrumental type. In this context, instrumental harm (e.g., using one person to achieve a goal, such as sacrificing an individual to save others) was judged harshly because process entailed a deliberate intention to cause harm, for a greater good. However, incidental harm (e.g., unintentionally causing harm while pursuing a goal) was more morally acceptable because it lacked the intention to harm, considering the similar consequences.

Another study by Cushman et al. (2006) focused on the distinction between intentional and unintentional harm, including the impact on moral judgment. Furthermore, incidental harms were judged less harshly compared to instrumental due to lack of malicious intent in related situations. Foot (2002) also discussed the tension between consequentialist and deontological approaches to moral reasoning, particularly in the context of instrumental and incidental harm. In utilitarian reasoning, which focused on outcomes, both instrumental and incidental harm were judged similarly, as long as the consequences were the same. However, deontologists tended to view instrumental harm as morally unacceptable because it entailed the use of others as means to an end, regardless of

the consequences. This explained why majority of the participants viewed instrumental harm as unacceptable, due to the exhibition of more deontological moral judgment.

Hypothesis 4 (H4) stated that self-involvement dilemmas was more acceptable compared to the others' involvement type. This was based on the assumption that individuals might be more forgiving or justifying certain actions when implicated. However, the results only partially supported the hypothesis, because both self-involvement and others' involvement dilemmas were rated at a moderate level of acceptability. These fell between incidental and instrumental dilemmas, with post hoc tests reporting no significant difference between the two conditions. The finding suggested that personal involvement, did not significantly affect moral acceptability judgments unless the intentional nature of the harm was clearly defined. The moderate ratings for these scenarios reflected moral ambivalence or shared responsibility, and it does not conclusively support the hypothesis. Therefore, H4 was weakly supported, resulting in the need for further exploration of the psychological mechanisms fundamental self-other distinctions in moral judgment.

In view of the description above, Kantian ethics, a classic deontological theory, stated that moral decisions were driven by duties and universal principles, rather than the consequences of actions. In this framework, certain actions were morally wrong regardless of whether the affect oneself or others. Moral law, stated that one must act in a universally acceptable manner, as well as respect the autonomy and dignity of all individuals. When confronted with moral dilemmas, deontologists such as Kant reported that the acceptance of harmful actions, violations of promises, or engaging in deceit were inherently wrong, irrespective of the outcomes. This viewpoint showed that moral rules were held universally, making certain actions equally unacceptable in scenarios concerning personal harm or detrimental to others (Bennett, 2015). The result differed from the study by Lotto et al. (2013), that self-involvement dilemmas were more accepted than the involvement of other people, specifically if the scenario was incidental. This showed the participants considered dilemmas, concerning themselves or other people, to be unacceptable. The finding was in line with the tendency of participants towards deontological moral decisions.

Asides from testing the core hypotheses, this study conducted exploratory analyses examining emotional arousal differences across the four dilemma types. Consistent with expectations, incidental harm dilemmas prompted significantly higher arousal than instrumental, self-involvement, or others' involvement scenarios. This pattern suggested that unintentional harm resulted in heightened emotional responses, due to the tension between ambiguous intent and negative outcomes, thereby generating moral conflict

even when no harm was intended. The finding was in line with the ETIC model (Emotional arousal–Theory of mind–Inhibitory Control) proposed by Buon et al. (2017) which extended dual process theory by reviewing how moral judgment unfolded when intention and outcome diverged. The ETIC stated that incidental harm scenarios generated strong affective responses (System 1). These were evaluated using the theory of mind and inhibitory control (System 2), to override initial aversive intuition. The inadequacy of this regulation due to high conflict or ambiguity, led to the elevation of subjective arousal. This was supported by Crockett et al. (2014) who stated personal and self-beneficial dilemmas, particularly in relation to incidental harm, produced higher levels of arousal and more negative valence, even when harm was unintentional. In these scenarios, participants faced moral tension between empathic intuition and rational justification, increasing emotional arousal regardless of intent.

Schwartz et al. (2022) also reported that incidental harm triggered stronger emotional reactions compared to instrumental, particularly among individuals with intuitive reasoning styles. The experiments showed that intentionality and harm severity collaboratively drove arousal and moral condemnation, specifically when harm was unintended but severe. The participants who engaged in deliberate reasoning were more forgiving of minor incidental harm, showing that reasoning style regulated emotional sensitivity to unintended harm. Additionally, these theoretical and empirical insights described why incidental harm dilemmas uniquely triggered intense emotional reactions in the selected sample. The convergence of ambiguous intent, perceived severity, and personal relevance appeared to strengthen the affective responses, consistent with an integrated cognitive-affective model of moral judgment.

The findings support the inference that moral cognition was sensitive to the structure of dilemmas, including individual differences in emotional responsiveness. Intention, causality, and personal involvement played major roles in shaping moral evaluations and emotional experiences. Although the gender variable was not a strong moderator of moral acceptability or judgment type, it influence arousal intensity, which partially accounted for differences in reasoning style. The findings of this study offered important theoretical implications for understanding the interplay between emotion, gender, and moral reasoning. It also supported the broader framework of dual process theory by showing that deontological judgments were preferred over the utilitarian type. Emotional arousal played a significant role in shaping these decisions. The partial support for gender-based differences in moral judgment and emotional arousal further refined the proposed model. This suggested that while both gender may not differ significantly in respective moral

acceptability ratings, women tended to experience stronger emotional arousal during morally complex scenarios. The increased arousal influenced the tendency towards deontological reasoning, consistent with the perspective that affective process drove intuitive moral responses. However, the absence of significant gender effects on moral acceptability also conformed with the similarities perspective, which implied moral cognition was more universally human than distinctly gendered. The substantial effects of dilemma type, specifically the distinct contrast between incidental and instrumental harm outlined the significance of perceived intentionality in moral cognition. The results showed that moral evaluation was a function of outcome assessment, as well as evolved from the dynamic interaction between emotional responses, cognitive control, and contextual features including causality and personal involvement. As a result, this study contributed to the increased analyses that moral decision-making was effectively understood through an integrative perspective, considering both cognitive-affective process and sociocultural influences, rather than simple categorical models based solely on gender or outcome ethics.

This study contained several critical omissions that might compromise the validity and generalizability of the findings. First, dilemma scenarios were adopted without any cultural adaptation. The lack of localization raised concerns about cultural bias, as participants did not completely understand the contextual nuances of dilemmas, and this potentially affected the authenticity of moral judgments. Second, the educational backgrounds of participants were not uniform. Differences in cognitive development and moral reasoning connected to educational level could act as a confounding variable, introducing variability unrelated to the experimental manipulations but still influenced the outcomes. Third, the scheduling of the experimental sessions was spread across morning, afternoon, and evening, without accounting for how time-related factors might influence participants' emotional and cognitive states. Additionally, the variations in alertness, mood, and fatigue could act as extraneous variables, affecting responses to emotionally charged moral dilemmas. These overlooked aspects suggested that the study might have underestimated the complexity of factors influencing moral decision-making, limiting the interpretive power and ecological validity of the inferences drawn.

Conclusions

In conclusion, this study investigated the effects of gender and dilemma types on moral judgment, acceptability, arousal, and valence. The results showed both genders preferred deontological decisions, with females exhibiting a non-significant tendency toward

stronger related responses. Additionally, the variable did not significantly influence moral acceptability ratings. The female participants reported significantly higher emotional arousal, and no gender differences was detected in respect to valence. Based on this context, incidental dilemmas prompted the highest arousal compared to other types, implying that unintentional but ambiguous harm provoked stronger emotional responses. Incidental dilemmas also outlined the importance of intentionality in moral judgment. Regarding the involvement scenario, self-involvement dilemmas were not significantly more acceptable compared to the others' involvement. Generally, these findings suggested that emotional arousal and dilemma structure, particularly intentionality and personal relevance, played a substantial role in moral decision-making than gender, supporting an integrated cognitive-affective perspective.

Suggestion

Future studies needs to consider several factors, such as adopt an organized time slot/ schedule for conducting experiments to avoid the materialization of time-difference variables as pollutants. Additionally, there is need to focus on inclusion and exclusion criteria for participants to obtain homogeneous responses. Cultural adaptations should be considered in using dilemma scenarios that are more contextual to daily activities and norm in Indonesia. Future studies should also explore additional emotional variables (e.g., guilt, empathy), use neurophysiological measures of arousal, or manipulate cultural framing to examine how moral cognition operated in diverse populations. Qualitative analyses helped described the narratives and justifications that accompanied different moral decisions, providing comprehensive insight into the psychological mechanisms at play.

The several limitations did not reduce the validity of the results from the experimental study. These results had practical implications as a reference/baseline for exploring the cognitive elements of moral decision and understanding the originating behaviors. The studies on moral decisions and dilemmas are still quite limited in Indonesia. Therefore, it is hoped that the present study provided additional scientific analysis through laboratory experimental evidence in the context of moral decision and judgments.

Funding

This study received grant from Institute for Research and Community Service, State Islam University Sunan Kalijaga Yogyakarta.

Acknowledgment

The is grateful to Awa Fauzia Malchan, Aji Wibowo Pamungkas, and Elok Jauharotus for their valuable assistance throughout the study process. The author is also grateful to all participants whose time and responses made this study possible.

References

Aurelia, M. Z., van Prooijen, J. W., & Van Lange, P. A. M. (2025). How do people morally judge corruption? A comparison between the Netherlands and Indonesia. *European Journal of Social Psychology*, *55*(3), 472–489. https://doi.org/https://doi.org/10.1002/ejsp.3130

Bechara, A. (2004). The role of emotion in decision-making: Evidence from neurological patients with orbitofrontal damage. *Brain and Cognition*, *55*(1), 30–40. https://doi.org/https://doi.org/10.1016/j.bandc.2003.04.001

Bennett, C. (2015). *What is this thing called ethics?* (2nd ed.). Routledge

Benhabib, S. (1985). The generalized and the concrete other: The Kohlberg-Gilligan controvers and feminist theory. *Praxis International*, *5*(4), 402-424. https://www.ceeol.com/search/journal-detail?id=615

Bianchin, M. & Angrilli, A. (2012). Gender differences in emotional responses: A psychophysiological study. *Physiology & Behavior, 105*(4), 925–932. https://doi.org/10.1016/j.physbeh.2011.10.031

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: the Self-Assessment Manikin and the Semantic Differential. *Journal of behavior therapy and experimental psychiatry, 25*(1), 49-59. https://doi.org/10.1016/0005-7916(94)90063-9

Buon, M., Habib, M., & Frey, D. (2017). Moral development: Conflicts and compromises. In J. A. Sommerville & J. Decety (Eds.), *Social cognition: Development across the life span* (pp. 129–150). Routledge/Taylor & Francis Group.

Buss, D. M. (2019). *Evolutionary psychology: The new science of the mind* (6th ed.). Pearson. https://batrachos.com/sites/default/files/pictures/1_5_HN_Books/Buss_2019_Evolutionary%20Psychology_The%20New%20Science%20of%20the%20Mind.pdf

Carmona-Perera, M., Marti, C., Pérez-García, M., & Verdejo-García, A. (2013). Valence of emotions and moral decision-making: Increased pleasantness to pleasant images and decreased unpleasantness to unpleasant images are associated with utilitarian choices in healthy adults. *Frontiers in Human Neuroscience*, *7*, 00626. https://doi.org/10.3389/fnhum.2013.00626

Conway, P., & Gawronski, B. (2013). Deontological and utilitarian inclinations in moral decision making: A process dissociation approach. *Journal of Personality and Social Psychology*. https://doi.org/https://doi.org/10.1037/a0031021

Crockett, M. J., Kurth-Nelson, Z., Siegel, J. Z., Dayan, P., & Dolan, R. J. (2014). Harm to others outweighs harm to self in moral decision making. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(48), 17320–17325. https://doi.org/10.1073/pnas.1408988111

Cushman, F. (2013). Action, outcome, and value: A dual- system framework for morality. *Personality and Social Psychology Review*, *17*(3), 273–292. https://doi.org/10.1177/1088868313495594

Cushman, F., Young, L., & Greene, J. D. (2010). Our multi-system moral psychology: Towards a consensus view. In J. M. Doris (Ed.), *Oxford handbook of moral psychology* (pp. 47-71). Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780199582143.003.0003

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*(12), 1082–1089. https://doi.org/10.1111/j.1467-9280.2006.01834.x

Domes, G., Schulze, L., Böttger, M., Grossmann, A., Hauenstein, K., Wirtz, P. H., Heinrichs, M., & Herpertz, S. C. (2010). The neural correlates of sex differences in emotional reactivity and emotion regulation. *Human Brain Mapping*, *31*(5), 758–769. https://doi.org/10.1002/hbm.20903

Doğruyol, B., Alper, S., & Yilmaz, O. (2019). The five-factor model of the moral foundations theory is stable across WEIRD and non-WEIRD cultures. *Personality and Individual Differences*, *151*, 109547. https://doi.org/https://doi.org/10.1016/j.paid.2019.109547

Eagly, A. H., & Wood, W. (2012). Social role theory. In P. A. M. Van Lange, A. W. Kruglanski, & E. T. Higgins (Eds.), *Handbook of theories of social psychology* (pp. 458–476). Sage Publications Ltd. https://doi.org/10.4135/9781446249222.n49

Ellemers, N., van der Toorn, J., Paunov, Y., & van Leeuwen, T. (2019). The psychology of morality: A review and analysis of empirical studies published from 1940 through 2017. *Personality and Social Psychology Review*, *23*(4), 332–366. https://doi.org/10.1177/1088868318811759

Foot, P. (2002). *Virtues and vices: And other essays in moral philosophy* (online ed.). Oxford Academic. https://doi.org/10.1093/0199252866.003.0011

Friesdorf, R., Conway, P., & Gawronski, B. (2015). Gender differences in responses to moral dilemmas: A process dissociation analysis. *Personality and Social Psychology Bulletin*, *41*(5), 696–713. https://doi.org/10.1177/0146167215575731

Fumagalli, M., Ferrucci, R., Mameli, F., Marceglia, S., Mrakic-Sposta, S., Zago, S., Lucchiari, C., Consonni, D., Nordio, F., Pravettoni, G., Cappa, S., & Priori, A. (2010). Gender-related differences in moral judgments. *Cognitive Processing*, *11*(3), 219–226. https://doi.org/10.1007/s10339-009-0335-2

Gamez-Djokic, M., & Molden, D. (2016). Beyond affective influences on deontological moral judgment: The role of motivations for prevention in the moral condemnation of harm. *Personality and Social Psychology Bulletin*, *42*(11), 1522–1537. https://doi.org/10.1177/0146167216665094

Gawronski, B., & Beer, J. S. (2017). What makes moral dilemma judgments "utilitarian" or "deontological"? *Social Neuroscience*, *12*(6), 626–632. https://doi.org/10.1080/17470919.2016.1248787

Gilligan, C. (1993). *In a Different Voice: Psychological Theory and Women's Development*. Harvard University Press. https://doi.org/10.2307/j.ctvjk2wr9

Gleichgerrcht, E., & Young, L. (2013). Low levels of empathic concern predict utilitarian moral judgment. *PloS One*, *8*(4), e60418. https://doi.org/10.1371/journal.pone.0060418

Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Personality and Social Psychology Bulletin*, *96*(5), 1029–1046. https://doi.org/10.1037/a0015141

Greene, J. D. (2014). The cognitive neuroscience of moral judgment and decision making. In M. S. Gazzaniga & G. R. Mangun (Eds.), *The cognitive neurosciences* (5th ed., pp. 1013–1023). Boston Review. https://doi.org/10.7551/mitpress/9504.003.0110

Greene, J. D. (2007). Why are VMPFC patients more utilitarian? A dual-process theory of moral judgment explains. *Trends in Cognitive Sciences*, *11*(8), 322–323. https://doi.org/10.1016/j.tics.2007.06.004

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*(2), 389–400. https://doi.org/10.1016/j.neuron.2004.09.027

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*(5537), 2105–2108. https://doi.org/10.1126/science.1062872

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Reviewa*, *108*(4), 814–834. https://doi.org/10.1037/0033-295X.108.4.814

Haidt, J., & Kasebir, S. (2010). Morality. In S. Fiske, D. Gilbert, & G. Lindzey (Eds.), *Handbook of social psychology* (5th ed., pp. 797-832). Wiley.

Harenski, C. L., Antonenko, O., Shane, M. S., & Kiehl, K. A. (2008). Gender differences in neural mechanisms underlying moral sensitivity. *Social Cognitive and Affective Neuroscience*, *3*(4), 313–321. https://doi.org/10.1093/scan/nsn026

Hauser, M. D., Cushman, F., Young, L., Jin, R., & Mikhail, J. (2007). A dissociation between moral judgment and moral action. *Proceedings of the National Academy of Sciences*, *104*(19), 8232-8237. https://doi.org/10.1111/j.1468-0017.2006.00297.x

Hofmann, W., Wisneski, D. C., Brandt, M. J., & Skitka, L. J. (2014). Morality in everyday life. *Science, 345*(6202), 1340–1343. https://doi.org/10.1126/science.1251560

Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist*, *60*(6), 581–592. https://doi.org/https://doi.org/10.1037/0003-066X.60.6.581

Ji, C. H. C., Ibrahim, Y., & Kim, S. D. (2009). Islamic personal religion and moral reasoning in social justice and equality: The evidence from indonesian college students. *The International Journal for the Psychology of Religion*, *19*(4), 259–274. https://doi.org/10.1080/10508610903143537

Jin, W. Y., & Peng, M. (2021). The effects of social perception on moral judgment. *Frontiers in Psychology*, *11*, 557216. https://doi.org/10.3389/fpsyg.2020.557216

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). Rand McNally.

Labvanced. (2025). *Professional experiments made easy* (Apparatus & software). Labvanced. https://www.labvanced.com/

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2008). International affective picture system (IAPS): Affective ratings of pictures and instruction manual. *Technical Report A-8*. University of Florida.

Lotto, L., Manfrinati, A., & Sarlo, M. (2013). A new set of moral dilemmas: norms for moral acceptability, decision times, and emotional salience. *Journal of Behavioral Decision Making*, *27*(1), 57–65. https://doi.org/https://doi.org/10.1002/bdm.1782

Ludwig, J., Reisenzein, R., & Hiemisch, A. (2020). Effects of instrumentality and personal force on deontological and utilitarian inclinations in harm-related moral dilemmas. *Frontiers in Psychology*, *11*, 1222. https://doi.org/https://doi.org/10.3389/fpsyg.2020.01222

McDonald, K., Graves, R., Yin, S., Weese, T., & Sinnott-Armstrong, W. (2021). Valence framing effects on moral judgments: A meta-analysis. *Cognition*, *212*, 104703. https://doi.org/https://doi.org/10.1016/j.cognition.2021.104703

McLean, C. P., & Anderson, E. R. (2009). Brave men and timid women? A review of the gender differences in fear and anxiety. *Clinical Psychology Review*, *29*(6), 496–505. https://doi.org/https://doi.org/10.1016/j.cpr.2009.05.003

McNair, S., Okan, Y., Hadjichristidis, C., & de Bruin, W. B. (2018). Age differences in moral judgment: Older adults are more deontological than younger adults. *Journal of Behavioral Decision Making*, *32*(1), 47–60. https://doi.org/https://doi.org/10.1002/bdm.2086

Oktrivina, A., Murni, Y., & Pramesti, A. (2024). Exploring ethical decision-making in accounting: The mediating role of moral judgment among professional accountants. *SEIKO: Journal of Management & Business*, *7*(2), 453–462. https://doi.org/https://journal.stieamkop.ac.id/index.php/seiko/article/view/7400/5420

Packer, D. J., Fujita, K., & Herman, A. D. (2021). Outgroup moral exclusion reduces reliance on moral principles. *Journal of Experimental Social Psychology, 93*, 104092. https://doi.org/10.1016/j.jesp.2020.104092

Qian, Y., Takimoto, Y., Wang, L., & Yasumura, A. (2024). Exploring cultural and gender differences in moral judgment: A cross-cultural study based on the CNI model. *Current Psychology*, *43*(6), 5243–5253. https://doi.org/10.1007/s12144-023-04662-6

Ridgeway, C. L. (2001). Gender, status, and leadership. *Journal of Social Issues*, *57*(4), 637–655. https://doi.org/https://doi.org/10.1111/0022-4537.00233

Rybash, J. M., Hoyer, W. J., & Roodin, P. A. (1983). Responses to moral dilemmas involving self versus others. *The International Journal of Aging and Human Development*, *18*(1), 73–77. https://doi.org/10.2190/ah2h-dfbx-2r13-07xq

Sachdeva, S., Iliev, R., Ekhtiari, H., & Dehghani, M. (2015). The role of self-sacrifice in moral dilemmas. *PLOS ONE*, *10*(6), e0127409. https://doi.org/10.1371/journal.pone.0127409

Schwartz, Flora, Djeriouat, Hakim, & Trémolière, Bastien. (2022). Judging accidental harm: Reasoning style modulates the weight of intention and harm severity. *Quarterly Journal of Experimental Psychology*, *75*(12), 2366–2381. https://doi.org/10.1177/17470218221089964

Urbaniak, G. C., & Plous, S. (2013). *Research randomizer* (Version 4.0) [Computer program]. http://www.randomizer.org/

Wei, Y. (2023). Gender differences in moral judgement and relevant factors. *Journal of Education, Humanities and Social Sciences*, *22*, 569-573. https://doi.org/10.54097/ehss.v22i.13061

Zhang, Q., & Thoma, S. (2016). An empirical cross-cultural study of moral judgment development in mainland China. *ETHICS IN PROGRESS*, *7*, 228–245. https://doi.org/10.14746/eip.2016.1.15

Received 16 Januari 2025
Revised 22 July 2025
Accepted 29 July 2025