Applied & Educational Psychology (2025)
Clausius Scientific Press, Canada

DOI: 10.23977/appep.2025.060306
ISSN 2523-5842 Vol. 6 Num. 3

Adolescents' Acceptance of Artificial Intelligence in Moral Dilemmas

Zhang Yiling

Department of Psychology, College of Education, Yanshan Campus, Guangxi Normal University, Guangxi, Guilin, China

Keywords: Artificial Intelligence, Moral Dilemma, Utilitarianism, Deontology

Abstract: As AI increasingly participates in social decision-making, adolescents' acceptance of AI decisions in moral dilemmas has become central to understanding human–machine ethics. Yet few studies have explored how adolescents respond to utilitarian versus deontological decisions made by different agents. This study used a between-subjects design with 180 middle school students, employing adapted versions of the trolley and footbridge dilemmas. Participants were randomly assigned to either an "AI decision" or "self-decision" condition, and acceptance was rated on a 6-point Likert scale. Results indicated a general adolescent preference for utilitarian outcomes. Furthermore, adolescents with more favorable attitudes toward AI showed greater acceptance of both types of AI decisions. By integrating utilitarian and deontological perspectives, this study offers insights into adolescent moral cognition in the age of intelligent technologies and underscores the need for ethically aligned AI system design.

1. Introduction

Artificial Intelligence (AI) is increasingly regarded as a moral agent capable of ethical decision-making [1]. Psychologically, its interaction with moral judgment involves complex cognitive, emotional, and behavioral mechanisms. With capabilities in perception, reasoning, and decision-making, AI can simulate and assist human moral reasoning via algorithms[2]. Classic dilemmas such as the trolley and footbridge cases are widely used to study this, highlighting two moral orientations: utilitarianism, which maximizes outcomes, and deontology, which emphasizes duties over consequences.

To explore adolescents' moral evaluations of AI, researchers have commonly adopted classic moral dilemmas such as the trolley dilemma and the footbridge dilemma as foundational paradigms[3]. The trolley dilemma presents a conflict scenario in which a runaway trolley is headed toward five innocent workers on the tracks. Participants must choose between actively diverting the trolley to another track, thereby saving five lives but sacrificing one person on the side track, or doing nothing and allowing the five to die. These dilemmas reflect two distinct ethical orientations: utilitarianism, which evaluates morality based on outcomes and seeks to maximize overall welfare[4], and deontology, which judges morality based on adherence to duty, rights, and inherent moral rules, regardless of consequences.

Building on these paradigms, a number of studies have examined how people evaluate AI's

decisions in moral dilemmas. Findings suggest that humans may apply different moral standards to AI. For example, people tend to find it more acceptable for AI to make utilitarian decisions, such as sacrificing one to save four[5]. Bonnefon and colleagues found that in the context of autonomous vehicles, participants were more tolerant of AI engaging in utilitarian actions and even evaluated such choices as more ethical than equivalent human actions. Longoni, using virtual reality technology, similarly found that people prefer AI to act in a utilitarian manner—favoring decisions that maximize the number of lives saved[6].

However, not all research aligns with this trend. Some studies indicate that adolescents may not necessarily prefer AI to adopt utilitarian strategies, nor do they always evaluate such decisions more positively than those made by humans. Moral judgments of AI are influenced by a host of contextual and cultural factors. For instance, in Japanese cultural contexts, the proportion of participants judging any action as morally wrong is relatively low, regardless of ethical orientation[7]. Similarly, the Swiss referendum system, which emphasizes "rule fairness," aligns more with deontological principles. These findings underscore that moral reasoning and behavior vary significantly not only between societies but also within them[8].

Most prior research centers on expectations and evaluations of AI behavior in Foot's scenarios, while few studies examine how adolescents judge AI versus their own decisions. Yet, AI and human agents exhibit differences in moral reasoning[9], and adolescence is a key stage of biological, psychological, and social development. Adolescents tend to view AI as more utilitarian[10]but prefer making decisions themselves in emotionally or socially significant situations[11]

AI moral decision acceptance has been studied across frameworks such as rational reliance, elaboration likelihood model (ELM), identity threat, dual-process theory, and cross-cultural models. According to ELM, AI acceptance depends on alignment between users' cognitive styles and system design, forming a motivation–cue fit[12]. Dual-process theory shows task type influences acceptance: utilitarian tasks (e.g., finance) increase AI adoption by 29% compared to hedonic tasks (e.g., art), moderated by cultural values [13]. A cross-cultural meta-analysis proposed a "technology–individual–culture" model emphasizing balance between universality and contextualization in AI design.

Moreover, research reveals that direct acceptance of AI decisions is generally lower than indirect acceptance, due to explicit social norms. For instance, when asked directly, participants weigh social preferences more heavily (68%), with institutional trust playing a significant moderating role. Social desirability biases also contribute to gender differences in AI .

This study explores whether adolescents differ in their acceptance of AI versus self-made decisions in moral dilemmas. Humans often elicit stronger empathic identification due to psychological proximity. Dual-process theory suggests individuals engage in more deliberate reasoning when evaluating AI decisions. In contrast, the "uncanny valley" effect[14] indicates human-like AI may cause discomfort, while the "threat to human distinctiveness" hypothesis posits that AI challenges uniquely human traits, prompting stricter judgment—such as in scenarios where AI autonomously sacrifices one to save many.

2. Method

2.1 Participants

Using G*Power 3.1 for an independent-samples t-test (effect size = 0.5, $\alpha = 0.05$, power = 0.8, two groups), the required sample size was calculated as 128. This study recruited 180 middle school students, all with normal cognitive functioning and no prior experience in moral judgment experiments.

2.2 Materials and Procedure

This study employed a between-subjects design with one independent variable: the identity of the decision-maker (AI decision vs. self-decision). The dependent variable was participants' acceptance level of the moral decision made.

Measurement Instruments:

1) Attitudes Towards AI Scale (ATTARI-12)

Developed by Stein et al. (2024) and published in Scientific Reports, the original English version was validated with both German and American samples.

2) Moral Dilemma Materials:

Two adapted moral dilemmas were used, based on:

- The Trolley Problem (Foot's dilemma)
- The Footbridge Dilemma

Each dilemma was presented in two versions:

- AI as the decision-maker
- The participant as the decision-maker

Participants were randomly assigned to either version. Example scenarios:

- Trolley Dilemma Scenario
- "Imagine a high-speed vehicle suddenly encountering five construction workers on the track. Braking is no longer possible, and the workers cannot escape. Unless the AI system (or you) changes direction, all five workers will be killed. However, there is one worker on the side track who would be killed if the direction is changed."
- Footbridge Dilemma Scenario
- "A high-speed vehicle is about to pass under a footbridge. Suddenly, five workers appear on the road. Braking is impossible, and unless the AI system (or you) activates a special mechanism, the workers will die. Activating the mechanism, however, will cause the death of one worker standing on the footbridge."

Participants in the AI decision group were asked to rate the AI's decision (e.g., to switch direction or activate the mechanism) on a 6-point Likert scale (0 = completely unacceptable to 6 = completely acceptable). Participants in the self-decision group were asked to indicate what decision they would make and rate their own decision using the same scale.

2.3 Results

Adolescents' Acceptance of AI Decisions in Moral Dilemmas

(1) Between-group differences: AI vs. Self Decision

Participants in the AI decision group (M = 4.43) reported significantly lower acceptance scores than those in the self-decision group (M = 5.96). A t-test revealed a significant difference:

$t(148) = -4.379$, Cohen's d = 2.121, indicating a large effect size.

(2) Within AI Decision Group: Utilitarian vs. Deontological Acceptance

Within the AI decision group, participants showed significantly higher acceptance of utilitarian decisions (M = 5.70) compared to deontological decisions (M = 3.16), $t(36) = 5.694$, with a large effect size (Cohen's d = 2.714).

(3) Within Self-Decision Group: Utilitarian vs. Deontological Acceptance

In the self-decision group, participants also showed higher acceptance of utilitarian decisions (M = 5.71) than deontological decisions (M = 3.43),$t(6) = 4.382$, with a large effect size (Cohen's d = 1.656).

(4) Correlation Between AI Attitude and AI Decision Acceptance

Among the 81 participants in the AI decision group, a significant positive correlation was observed

between participants' attitudes toward AI (as measured by ATTARI-12) and their acceptance of AI decisions,r = 0.392. This suggests that the more positively adolescents evaluated AI (e.g., reliability, friendliness, usefulness), the more likely they were to accept its moral decisions, whether utilitarian or deontological.

3. Discussion

This study examined adolescents' acceptance of AI decisions in moral dilemmas, compared to decisions made by themselves. Results showed significantly lower acceptance of AI decisions (M = 4.43) than self-decisions (M = 5.96; t = -4.379) [15], likely due to reduced emotional involvement and psychological distance [16]. In line with dual-process theory, AI is viewed as a rational but intuition-free agent [17].

Adolescents favored utilitarian decisions over deontological ones, both for AI (M = 5.70 vs. 3.16) and self-decisions (M = 5.71 vs. 3.43). This supports the idea that utilitarian choices reinforce a prosocial self-image in collectivist cultures. As adolescents move toward post-conventional moral reasoning, they increasingly prioritize outcomes over rules. Moral disengagement may also reduce self-blame for utilitarian decisions, especially among those with higher autonomy[18].

A positive correlation between attitudes toward AI and acceptance of its decisions (r = 0.392, p < .001) aligns with the Elaboration Likelihood Model, suggesting transparent reasoning increases AI trust [19].

Limitations include the narrow age range (12–15), simplified dilemma types, and absence of anthropomorphic features in the AI. Future studies should explore age-related trends and cross-cultural differences in AI moral acceptance.

4. Conclusion

(1) Adolescents are more likely to accept moral decisions made by themselves than by AI.

(2) When AI makes the decision, adolescents prefer outcome-based utilitarian choices over rule-based deontological ones.

(3) When deciding themselves, adolescents favor utilitarian actions that sacrifice a few to save many.

(4) More positive attitudes toward AI lead to higher acceptance of its moral decisions, regardless of their ethical orientation.

References

[1] Zacharus Gudmunsen.(2024).The moral decision machine: a challenge for artificial moral agency based on moral deference.AI and Ethics,5(2),1-13.

[2] Gordon, E. C., Cheung, K., et al. (2024). "Moral enhancement and cheapened achievement: Psychedelics, virtual reality and AI." Bioethics 39 (3): 276-287.

[3] Jiang, Y., Zhang, W., Wan, Y., Gummerum, M., & Zhu, L. (2024). Adolescents are more utilitarian than adults in group moral decision-making. *Psych Journal, 14(2), 179-190. https://doi.org/10.1002/pchj.821*

[4] Caprara, G. V., Alessandri, G., & Bornstein, M. H. (2021). Individual and environmental correlates of adolescents' moral decision-making in moral dilemmas. Frontiers in Psychology, *12:8651977. https://doi.org/10. 3389/fpsyg.2021. 8651977*

[5] Bigman, Y. E., Gray, K., & Malle, B. F. (2022). Artificial intelligence and moral dilemmas: Perception of ethical decision-making in AI. Journal of Experimental Social Psychology*, 99, 104270. https://doi.org/10.1016/j.jesp. 2022. 104270*

[6] Kim, J., Giroux, M., & Lee, J. C. (2021). When do you trust AI? The effect of number presentation detail on consumer trust and acceptance of AI recommendations. Psychology & Marketing, 38(7)

[7] Luna Tacci, A., Manzi, F., Di Dio, C., Marchetti, A., Riva, G., & Massaro, D. (2023). Moral Context Matters: A study of Adolescents' Moral Judgment towards Robots.*Conference proceedings / book chapter.*

[8] Bentahila, L., & Fontaine, R. (2021). Universality and Cultural Diversity in Moral Reasoning and Judgment. Frontiers in Psychology, 12:764360. https://doi.org/10.3389/fpsyg.2021.764360
[9] Zhang, Z., Chen, Z., et al. (2022). "Artificial intelligence and moral dilemmas: Perception of ethical decision-making in AI." Journal of Experimental Social Psychology 101 : 104327.
[10] Ismatullaev Ulugbek Vahobjon Ugli & Kim SangHo.(2022).Review of the Factors Affecting Acceptance of AI-Infused Systems. Human factors,66(1),187208211064707-187208211064707. Zhang, Y., Wu, J., Yu, F., & Xu, L. (2023). Moral judgments of human vs. AI agents in moral dilemmas. Behavioral Sciences, 13(2), 181. https: //doi. org/10. 3390/bs13020181
[11] Irfan Ali, Nosheen Fatima Warraich & Khadijah Butt.(2025).Acceptance and use of artificial intelligence and AI-based applications in education: A meta-analysis and future direction. Information Development,41(3),859-874.
[12] Jiang, P., Niu, W., Wang, Q., Yuan, R., & Chen, K. (2024). Understanding users' acceptance of artificial intelligence applications: A literature review. Behavioral Sciences, 14*(8), Article 671. https://doi.org/10.3390/bs14080671*
[13] Laakasuo Michael.(2023).Moral Uncanny Valley revisited – how human expectations of robot morality based on robot appearance moderate the perceived morality of robot decisions in high conflict moral dilemmas .Frontiers in Psychology,14.
[14] Zhang, Y. W. J. Y. F. &. X. L. (2023). Moral judgments of human vs. AI agents in moral dilemmas. Behavioral Sciences, 13(2), 181.
[15] Cian, C. L. (2020). "Artificial Intelligence in Utilitarian vs. Hedonic Contexts: The "Word-of-Machine" Effect." Journal of Marketing 86 (1): 91-108.
[16] Lydia Cao,Sara Hennessy & Rupert Wegerif.(2025).Thinking fast, slow, and ahead: Enhancing in-service teacher contingent responsiveness in science discussion with mixed-reality simulation. Computers & Education,237,105389-105389.
[17] Sedikides,C.,Gaertner,L.,& Toguchi,Y. (2003). Pancultural self-enhancement. Journal of Personality and Social Psychology, 84*(1), 60–79. https://doi.org/10.1037/0022-3514.84.1.60*
[18] Caro-Burnett, J. and Kaneko, S. (2022). "Is Society Ready for AI Ethical Decision Making? Lessons from a Study on Autonomous Cars." Journal of Behavioral and Experimental Economics 98 : 101881.