Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Intentionality attribution in trolley-type moral dilemmas across the United States, Japan, and China: a preregistered cross-cultural study

Zhiheng Han

s2120426@jaist.ac.jp

Japan Advanced Institute of Science and Technology

Masaharu Mizumoto

Japan Advanced Institute of Science and Technology

Article

Keywords:

Posted Date: November 1st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10654027/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Intentionality attribution plays an important role in moral cognition, but it remains unclear whether the factors shaping such judgements operate similarly across cultural contexts. This study examined the effects of intent, personal force, and actor–observer perspective on intentionality attribution in trolley-type moral dilemmas across the United States, Japan, and China. In this preregistered between-participants study, 611 participants from China, Japan, and the United States were assigned to conditions in a 2 × 2 × 2 × 3 design manipulating intent, personal force, perspective, and region. Binary responses concerning moral choice, moral acceptability, and intentionality attribution were analysed using logistic regression models. The preregistered tests examined Intent × Region and Perspective × Region interactions and the main effect of Personal Force. Post hoc simulation-based sensitivity analyses evaluated the interaction magnitudes detectable with approximately 80% power under hypothesis-consistent directional patterns. The preregistered cultural interaction hypotheses were not supported. Intentionality attribution was more likely in means than side-effect conditions and in first-person than third-person conditions in pooled analyses, whereas Personal Force did not significantly predict intentionality attribution. Baseline intentionality attribution rates differed substantially across regions. The sensitivity analyses indicated that the study was primarily capable of detecting relatively large cultural moderation effects under the assumed interaction patterns. Secondary, non-preregistered analyses suggested possible differences between the predictor patterns of moral judgement and intentionality attribution, but these findings were exploratory. The study did not detect the preregistered cultural moderation effects. The pooled results identify intent and perspective as predictors of intentionality attribution in the present samples, but the nonsignificant interactions should not be interpreted as evidence of cultural equivalence. Further preregistered research should use designs powered to detect smaller interaction effects and experimentally vary question order.

Introduction

If a person faces a situation in which any available choice leads to a foreseeable but inevitable bad outcome and chooses an action, did that person act intentionally? What if the person were you?

Many studies in moral psychology have shown that people from different cultures often reach strikingly different moral judgements in the same situations. In classic moral dilemmas such as the trolley problem, participants from different societies vary widely in their choices and judgements. However, morality is only one component of how people evaluate actions. Another central aspect of moral cognition about action concerns intentionality attribution, that is, whether an agent is seen as having intentionally brought about a harmful outcome.

A fundamental question that arises is, then: do cultural differences in moral judgement extend to judgements about intentional action, or do people rely on a more universal framework when deciding whether an action was intentional? The present study addresses this question by examining intentionality attribution in trolley-type moral dilemmas across samples from the United States, Japan,

and China. These samples were selected to provide contrasting cultural and geographic contexts (Muthukrishna et al. 2020).

In moral psychology, intentionality has long been regarded as central to moral judgement. Classic work on the folk concept of intentionality argues that people rely on a shared structure when deciding whether an action is intentional, typically involving components such as desire, belief, intention, awareness, and skill (Malle and Knobe 1997). From this perspective, judgements of responsibility and blame are thought to build on prior assessments of whether the behaviour was intentional (Malle 2006b). However, Knobe (2003) challenged this traditional, one-directional view by showing that moral evaluation can also influence how people judge intentionality. In his famous "chairman" vignettes, people tended to say that a harmful side effect was brought about intentionally, but a helpful one was unintentional—even though the agent's mental states were held constant (not caring about the side effect at all). This moral asymmetry, now known as the Knobe effect or side-effect effect, suggests that moral evaluation and intentionality attribution are deeply intertwined. Subsequent studies found similar patterns in other populations, including children and clinical groups (Leslie et al. 2006), and in translations of the task into other languages, such as Hindi (Knobe and Burra 2006), leading some authors to propose that this asymmetry reflects a robust feature of how people think about minds and morality (Sauer 2014). More recent work, however, suggests cultural differences in this effect (Robbins et al. 2017), and therefore, the extent of the possible cultural variance and its cause deserve further exploration.

A related line of research examines how intentionality and moral judgement interact in moral dilemma scenarios. Previous research in Western samples has shown a strong difference between harm-as-means and harm-as-side-effect, where harm-as-means was taken to be a case in which harm was intended. Dual-process models of moral cognition suggest that factors like personal force (e.g., pushing a person as opposed to just flipping a switch) and the agent's intent (means vs. side effect) shape both emotional reactions and moral evaluations (Greene et al. 2009). Large cross-cultural studies have replicated these effects in Western samples but found more mixed patterns globally. Bago et al. (2022) found that people in Western countries judged harm used as a means as much less acceptable than harm that happened as a side effect. Interestingly, however, Bago and colleagues did not observe this difference in East Asian samples, suggesting that the distinction might not carry the same meaning for all cultures. Thus, the pattern of intentionality attributions (whether someone acted intentionally) in moral dilemmas might also differ culturally. Yet, much of this work has focused on moral acceptability rather than explicit judgements of intentionality, and few studies have systematically combined moral dilemma scenarios with direct measures of intentionality across cultures.

In addition, although the Knobe effect has been formulated from a third-person perspective, there has been a long discussion of how perspective influences social and moral thinking (Jones and Nisbett 1972; Malle 2006a; Malle and Knobe 1997). This framework has been extended to evaluations of actions in moral dilemmas, which has shown that actors are more likely than observers to say that choosing an action (rather than an omission) is morally less permissible (Jiang et al. 2022; Miller et al. 2014; Nadelhoffer and Feltz 2008). At the same time, cultural differences in moral choice and judgement in

moral dilemmas have also been well-documented (Awad et al. 2020; Gold et al. 2014; Qian et al. 2024), where East Asians (in particular, Chinese and Japanese) have shown stronger tendencies for action aversion (omission). We may therefore expect the corresponding cultural differences in the perspective effect concerning intentionality attribution, too. Yet few studies have investigated whether the same asymmetry and cultural variation are observed in intentionality attribution.

To address these questions, the present study examines intentionality attribution in moral dilemmas across samples from the United States, Japan, and China. We compare intentionality attribution with moral evaluation to test whether the observed patterns vary across the three samples.

As we will show, the results revealed similar observed effects of intent and perspective across the three samples, alongside substantial cultural variation in baseline response rates. Because the cultural interaction tests were not designed to establish equivalence, these findings are treated as suggestive of shared features rather than proof of cross-cultural invariance.

Theoretical Framework and Hypotheses

Our study on intentionality attribution investigates the possible influence of the following three factors and their possible interactions with culture:

(1) the means–side effect distinction,

(2) the role of personal force, and

(3) the perspective asymmetry.

The means/side-effect distinction concerns whether harm is instrumental to achieving a goal or merely a by-product. Bago et al. (2022) found that Western participants judged harm-as-means as less morally acceptable than harm-as-side-effect, whereas this difference was weaker or absent in East Asian samples. Moral evaluation and intentionality attribution are related but conceptually distinct: a side effect can still be judged intentional (Knobe 2003), and cross-cultural variation has been reported in judgements of intentional action (Knobe and Burra 2006; Robbins et al. 2017). We therefore tested the limited prediction that culture may moderate the use of the means/side-effect distinction when participants explicitly judge intentionality. This prediction extends prior moral-judgement findings and should not be read as an established cultural difference in intentionality attribution. The second factor, personal force, concerns whether the action involves direct physical contact — for example, pushing a person — or indirect intervention, such as flipping a switch. Earlier research has shown that moral and emotional processing differ dramatically between these two types of actions. We examine whether personal force also increases perceived intentionality.

Finally, perspective asymmetry concerns a possible difference between judging one's own action and judging another person's action. Prior work reports actor–observer differences in attribution and, in

some moral-dilemma studies, lower permissibility judgements from the actor perspective (Jones and Nisbett 1972; Malle 2006a; Nadelhoffer and Feltz 2008). Cross-cultural research on causal attribution suggests that East Asian participants tend, on average, to assign relatively greater weight to situational constraints and contextual information, whereas Western participants often place greater emphasis on dispositional characteristics when explaining behaviour (Choi et al. 1999; Miyamoto and Kitayama 2002; Morris and Peng 1994). This contrast is not absolute: dispositional reasoning occurs across cultures, and cultural differences appear to depend partly on the salience of situational constraints and the perceived diagnosticity of the observed behaviour. Related research has also shown greater context sensitivity in attention and memory among Japanese than American participants (Masuda and Nisbett 2001). Because intentionality attribution involves interpreting an agent's mental states in relation to the circumstances surrounding an action, these differences in attributional focus provide a theoretical basis for examining whether the effect of actor versus observer perspective varies across cultural contexts. However, this connection is indirect, and direct evidence concerning cultural moderation of perspective effects in intentionality attribution remains limited. We therefore tested whether culture moderates the perspective effect on intentionality attribution. Our directional expectation was that the actor–observer contrast would be larger in the U.S. sample than in the Japanese and Chinese samples, but this prediction was necessarily tentative because previous studies had not directly tested the same interaction for intentionality judgements.

Based on these frameworks, we formulated three hypotheses. The following preregistered hypotheses are reproduced here without substantive modification:

H1. (non-directional): Unlike Western participants, East Asian participants may not use the distinction between means and side effects as a basis for determining whether an action is intentional.

H2. (directional): The exertion of personal force tends to increase the perception that the agent intentionally sacrificed another person.

H3. (directional): The perspective asymmetry in intentionality attribution is expected to be stronger in Western cultures than in East Asian cultures.

Methods

Experimental Design

For this preregistered study, we employed a 2 (Perspective) × 2 (Personal Force) × 2 (Intent) across 3 countries (China, Japan, U.S.) design, manipulating four factors. This study was preregistered (Han and Mizumoto 2026a).

1. Intent (Means/Side Effect)

2. Personal Force (switch vs. push)

3. Perspective (actor vs. observer)

4. Region (Japan, China, U.S.)

We planned to recruit approximately 600 participants—about 200 for each of three target samples—through Credamo, Lancers, and Prolific Academic for the China-, Japan-, and U.S.-associated samples, respectively. The questionnaires were administered online using Google Forms. The study was designed, coordinated, administered, and managed by the research team at the Japan Advanced Institute of Science and Technology (JAIST) in Japan. No local investigators, collaborating research institutions, or physical research sites were involved in China or the United States. Participants in each target sample were randomly assigned to one of eight experimental conditions. The scenarios were based on Greene et al. (2009), and the study materials are publicly available (Han and Mizumoto 2026b).

Before data collection, we conducted an a priori power analysis in G*Power 3.1 for binary logistic regression using a two-sided Wald test, alpha = .05, power = .80, an anticipated odds ratio of 1.68 for the focal predictor, and R-squared with other predictors set to 0. The most parameterized planned models contained seven regression coefficients and yielded a minimum total sample size of 419. Because the complete between-participants design comprised 24 cells, we increased the recruitment target to approximately 600 participants, corresponding to about 25 participants per cell. The final analytic sample of 611 exceeded both the model-based minimum and the planned recruitment target. This original calculation was based on a focal logistic-regression coefficient rather than on a fully specified interaction shape; it therefore should not be interpreted as guaranteeing adequate power for small Region by Intent or Region by Perspective interactions.

After data collection, exclusion criteria were applied to ensure data quality. Participants who failed the attention check or did not report the survey language as their native language were excluded from the analysis. As a result, the final sample sizes were as follows: China (N = 219), Japan (N = 204), and the United States (N = 188), yielding a total sample of 611 participants. In total, 75 participants were excluded (China: 21, Japan: 22, U.S.: 32).

Variables

We measured the following three dependent variables:

Moral Choice (Q1): Would you/Should Joe hit the switch in this situation?

Moral evaluation (Q2): Do you think it is morally acceptable (for Joe) to hit the switch in this situation to avoid the deaths of five workers, but instead cause the death of one worker?

Intentionality Attribution (Q3): Suppose that, as a result, you/Joe did hit the switch. Do you think you/Joe intentionally sacrificed someone?

Q3 (the intentionality judgement) served as our primary dependent variable for hypothesis testing.

Question order was fixed as Q1 (choice), Q2 (moral acceptability), and Q3 (intentionality). This ordering was retained for procedural comparability with the source moral-dilemma paradigm and to ensure that participants first considered the action and its consequences before answering the intentionality question. We do not assume that this is the only or psychologically natural sequence of moral cognition. Because order was not randomised or experimentally manipulated, associations among Q1, Q2, and Q3 may partly reflect carryover or consistency effects and are interpreted only exploratorily.

Analysis Plan

Data were analysed using logistic regression models. The preregistered confirmatory tests were the two interaction tests and one main-effect test listed below:

H1: Intent (means vs. side effect) × Region interaction

H2: Main effect of Personal Force (push vs. switch)

H3: Perspective × Region interaction

The two remaining experimental variables were included as covariates in each hypothesis-specific model. Analyses of demographic variables, associations among Q1-Q3, moral-judgement outcomes, and joint response-type models were not preregistered and are reported only as secondary exploratory analyses.

Responses and predictor variables were dummy-coded into binary variables (0/1). Intent was coded as 1 = means condition and 0 = side-effect condition. Perspective was coded as 1 = first-person and 0 = third-person. Personal Force was coded as 1 = direct force (push) and 0 = indirect force (switch). Region was treated as a categorical variable with Japan as the reference group, with two dummy variables representing China and the United States.

Interaction-Specific Sensitivity Analysis

Because the original G*Power calculation was based on a focal logistic-regression coefficient rather than the specific interaction structures tested in H1 and H3, we conducted post hoc simulation-based sensitivity analyses. The simulations preserved the observed sample size, regional distribution, and allocation across the 24 experimental cells. Nuisance parameters other than the focal Intent coefficients were based on the pooled main-effects model. For H1, the Intent effect was set to zero in the Japanese and Chinese samples, while a positive Intent effect of varying magnitude was specified for the U.S. sample. For H3, the Perspective effect in the Japanese and Chinese samples was fixed at the pooled estimate, while an additional positive Perspective effect in the U.S. sample was varied. For each candidate interaction coefficient, 20,000 binary datasets were generated. The interaction model was compared with the corresponding main-effects model using a two-degree-of-freedom likelihood-ratio test at alpha = .05. The interaction coefficient yielding approximately 80% estimated power was treated as the minimum detectable interaction magnitude under the specified pattern. Because the results

depend on these assumed interaction shapes and nuisance parameters, they are interpreted as conditional sensitivity estimates rather than universal power values.

Results

We initially collected data from 686 participants across the three countries. After applying the exclusion criteria, the final sample consisted of 611 participants: 219 from China, 204 from Japan, and 188 from the United States.

The gender distribution was 74 males, 144 females, and one participant who did not report gender in China; 136 males and 68 females in Japan; and 98 males, 87 females, and three participants who did not report gender in the United States. The mean age was 32.0 years in China (standard deviation [SD] = 10.2), 45.0 years in Japan (SD = 9.9), and 43.1 years in the United States (SD = 12.9).

Overall Results

First, preliminary comparisons across regions regarding responses to each question showed significant differences. Table 1 summarizes the overall “yes” rates across the three countries.

Table 1. Overall summary of “yes” response rates

	China	Japan	U.S.
Choice (action) (Q1)	66.7%	30.4%	72.9%
Judgment (acceptability) (Q2)	48.9%	27.9%	69.7%
Intentionality (Q3)	31.5%	72.5%	79.8%

There were significant cross-cultural differences in moral choice (Q1) and moral evaluation (Q2). Japanese participants showed markedly lower rates of action choice and moral acceptability than participants from China and the United States. These differences were statistically significant (Q1: $\chi^2(2) = 86.27$, $p < .001$, Cramer's $V = .376$; Q2: $\chi^2(2) = 68.31$, $p < .001$, Cramer's $V = .334$).

A significant cross-cultural difference was also observed in intentionality attribution (Q3), although the pattern was markedly different from that of moral judgements (Q1 and Q2): Chinese participants showed much lower rates of intentionality attribution than Japanese and American participants did. This difference was also statistically significant (Q3: $\chi^2(2) = 118.20$, $p < .001$, Cramer's $V = .440$).

The breakdown of responses to Q3 is summarised in Figure 1.

As illustrated in Figure 1, intentionality attribution tended to be higher in means conditions than in side-effect conditions, and higher in first-person than in third-person perspectives across the three samples. Personal force appeared to have a relatively smaller effect compared to other factors.

Overall, intentionality attribution rates were relatively high among American and Japanese participants, even in side-effect conditions, consistent with patterns often discussed in relation to the Knobe effect.

As preliminary analyses, we conducted separate logistic regression analyses for each region, predicting intentionality attribution from intent type (means vs. side effect), personal force, and perspective. Across all three regions, Perspective significantly predicted intentionality attribution (Japan: $p = .002$; China: $p = .032$; U.S.: $p = .016$). Personal Force was not significant in any region. Intent did not significantly predict intentionality attribution in any region.

Supplementary Pooled Main-Effects Model

To examine the effects of Personal Force, Intent (means vs. side effect), Perspective, and Region on intentionality attribution (Q3), we conducted a logistic regression analysis using the cleaned dataset. The model is summarised in Table 2 and can be expressed as follows:

$$\text{logit}(\pi) = \alpha + \beta_1 D_{_intent} + \beta_2 D_{_perspective} + \beta_3 D_{_force} + \beta_4 D_{_china} + \beta_5 D_{_us}$$

In Tables 2–4, *b* denotes the regression coefficient, SE denotes the standard error, and OR denotes the odds ratio.

Table 2. Logistic Regression Predicting Intentionality Attribution: Main Effects of Intent, Perspective, Personal Force, and Region

Predictor	b	SE	z	p	OR
Intercept	0.491	0.227	2.166	0.030	1.63
Intent	0.482	0.189	2.551	0.011	1.62
Force	-0.226	0.190	-1.192	0.233	0.80
Perspective	0.840	0.192	4.374	< .001	2.32
China vs. Japan	-1.834	0.222	-8.253	< .001	0.16
U.S. vs. Japan	0.400	0.245	1.632	0.103	1.49

The results revealed significant effects of Intent, Perspective, and Region, whereas Personal Force was not significant. Intent significantly predicted intentionality attribution ($b = 0.48$, $p = .011$), indicating that means conditions were more likely to be judged as intentional than side-effect conditions.

Perspective showed a strong significant effect ($b = 0.84$, $p < .001$), with first-person scenarios more likely to be judged as intentional than third-person scenarios. Personal Force did not significantly predict intentionality attribution ($b = -0.23$, $p = .233$).

Regarding cultural differences, Chinese participants were significantly less likely than Japanese participants to judge actions as intentional (b = -1.83, p < .001), whereas the difference between U.S. and Japanese participants was not statistically significant (b = 0.40, p = .103).

In this pooled model, Intent and Perspective predicted intentionality attribution, whereas Personal Force did not. These main effects describe the overall sample but do not constitute support for H1 or H3, which concerned cultural moderation. Baseline regional differences were also substantial, particularly between China and Japan.

H1 Results — Means vs. Side Effect Across Cultures

To test whether cultural differences modulate the means/side-effect distinction (H1), we conducted a logistic regression analysis including Region × Intent interaction terms.

The model was specified as:

$$\text{logit}(\pi) = \alpha + \beta_1 D_{_intent} + \beta_2 D_{_china} + \beta_3 D_{_us} + \beta_4 D_{_perspective} + \beta_5 D_{_force} + \beta_6 (D_{_china} \times D_{_intent}) + \beta_7 (D_{_us} \times D_{_intent})$$

Table 3. Logistic Regression with Region × Intent Interactions (H1)

Predictor	b	SE	z	p	OR
Intercept	0.512	0.261	1.964	0.050	1.67
Intent	0.439	0.322	1.365	0.172	1.55
China vs. Japan	-1.825	0.311	-5.861	< .001	0.16
U.S. vs. Japan	0.315	0.331	0.952	0.341	1.37
Force	-0.226	0.190	-1.192	0.233	0.80
Perspective	0.840	0.192	4.378	< .001	2.32
Intent × China	-0.015	0.439	-0.034	0.973	0.99
Intent × U.S.	0.192	0.494	0.388	0.698	1.21

The results showed that none of the Region × Intent interaction terms were significant (ps > .69). A likelihood ratio test comparing models with and without the interaction terms further confirmed that including these interactions did not significantly improve model fit, $\Delta\chi^2(2) = 0.21$, $p = .90$.

Taken together, the analyses did not detect cultural moderation of the means/side-effect effect, and H1 was not supported in this sample. Because the original sample-size calculation was not tailored to a prespecified interaction shape, this null result should not be interpreted as establishing that the effect is identical across cultures. The simulation-based sensitivity analysis reported below further indicated that the present design was primarily sensitive to relatively large interactions under the assumed pattern.

H2 Results — Effect of Personal Force

Regarding whether exerting personal force increases perceived intentionality (H2), a logistic regression analysis with the basic model showed no significant effect of Personal Force on intentionality attribution (Table 2). This was also true for all region-specific models. Participants did not perceive actions involving direct physical force (pushing) as more intentional than those involving indirect intervention (flipping a switch).

To further assess the robustness of this finding, additional analyses were conducted using moral choice (Q1) and moral evaluation (Q2) as dependent variables. In contrast to the intentionality attribution model, Personal Force significantly predicted both moral choice (Q1: b = -0.457, p = .011) and moral evaluation (Q2: b = -0.656, p < .001). This pattern suggests that personal force plays a role in moral judgement, but not in judgements of intentionality.

Taken together, these results suggest that the effect of personal force observed in moral evaluation does not extend to intentionality attribution. Thus, H2 was not supported.

H3 Results — Perspective Asymmetry Across Cultures

To examine whether perspective asymmetry varies across cultures (H3), we conducted a logistic regression analysis including Region × Perspective interaction terms.

The model was specified as:

$$\text{logit}(\pi) = \alpha + \beta_1 D_{\text{–perspective}} + \beta_2 D_{\text{–china}} + \beta_3 D_{\text{–us}} + \beta_4 D_{\text{–intent}} + \beta_5 D_{\text{–force}} + \beta_6 (D_{\text{–perspective}} \times D_{\text{–china}}) + \beta_7 (D_{\text{–perspective}} \times D_{\text{–us}})$$

Table 4. Logistic Regression with Region × Perspective Interactions (H3)

Predictor	b	SE	z	p	OR
Intercept	0.417	0.248	1.682	0.093	1.52
Perspective	1.030	0.337	3.061	0.002	2.80
China vs. Japan	-1.653	0.300	-5.518	< .001	0.19
U.S. vs. Japan	0.444	0.312	1.423	0.155	1.56
Intent	0.483	0.189	2.553	0.011	1.62
Force	-0.222	0.190	-1.172	0.241	0.80
Perspective × China	-0.389	0.449	-0.865	0.387	0.68
Perspective × U.S.	-0.119	0.509	-0.235	0.814	0.89

The results revealed a significant main effect of Perspective, b = 1.03, SE = 0.34, z = 3.06, p = .002, OR = 2.80, indicating that first-person actions were more likely to be judged as intentional than third-person actions. However, neither the Perspective × China interaction, b = −0.39, p = .387, nor the Perspective × U.S. interaction, b = −0.12, p = .814, was statistically significant.

A likelihood-ratio test comparing models with and without the Region × Perspective interaction terms showed that adding the interactions did not significantly improve model fit ($\Delta\chi^2(2) = 0.81$, p = .67).

The analyses therefore did not provide evidence that the perspective effect differed across the three samples, and H3 was not supported. Although an overall perspective effect was observed, the nonsignificant interactions should not be interpreted as establishing cross-cultural equivalence.

Interaction-Specific Sensitivity Results

Under the assumed H1 pattern, the present design yielded approximately 80% estimated power at a Region × Intent interaction coefficient of 1.52, corresponding to an interaction odds ratio of 4.57. Under the assumed H3 pattern, the present design yielded approximately 80% estimated power at an additional U.S. Region × Perspective interaction coefficient of 1.86, corresponding to an interaction odds ratio of 6.42. These estimates indicate that the study was primarily sensitive to relatively large cultural moderation effects under the specified patterns; smaller or moderate interactions may have gone undetected.

Exploratory and Post Hoc Analyses

To further characterise the observed patterns, we conducted additional analyses examining demographic effects, relationships between judgement types, and the robustness of previously reported effects in moral dilemma research.

Demographic Effects: Age and gender were examined in exploratory analyses. To ensure data quality, one implausible age value (255 years old) was excluded from these demographic analyses only. Exploratory analyses indicated that age showed a small but significant effect on intentionality judgements (b = -0.022, OR = 0.98, p = .012), whereas gender did not show a significant effect (b = -0.35, OR = 0.70, p = .067). Increasing age was associated with a slightly lower likelihood of judging the agent's action as intentional. A similar (but non-significant) tendency was observed for female participants. Neither the Age × Region nor the Gender × Region interaction was statistically significant. These exploratory analyses therefore did not detect regional moderation of the demographic associations, but the nonsignificant interactions should not be interpreted as evidence of equivalence.

Associations Between Judgement Types: We next examined how moral responses relate to intentionality attribution. Logistic regression analyses showed that in separate bivariate models, both moral choice (Q1) and moral evaluation (Q2) were associated with intentionality attribution (Q3) (moral choice: p < .001, moral evaluation: p < .001). However, when Q1 and Q2 were entered simultaneously, only moral choice remained significant (p < .001), whereas moral evaluation did not (p = .905). Participants who

chose to perform the action (rather than omission) were significantly less likely to judge the agent's behaviour as intentional. This pattern indicates a systematic association between action endorsement and intentionality attribution, potentially reflecting differences in how participants construe agency or deliberateness in the scenario. In contrast, moral evaluation (Q2) did not show a consistent association with intentionality attribution.

Robustness and Replication of Moral Judgement Effects: As further post hoc analyses, to examine whether earlier studies on moral dilemmas are replicated in our study, we conducted logistic regression analyses with moral choice (Q1) and moral evaluation (Q2) as dependent variables.

For choice, personal force and region were significant predictors of moral choice. Participants were less likely to choose action (rather than omission) in scenarios involving personal force ($p = .011$). In addition, Japanese participants showed a significantly lower choice rate of action than Chinese and American participants (both $p < .001$). Intent and perspective were not significant predictors.

A similar pattern was observed for moral evaluation (acceptability). Personal force and region significantly predicted moral evaluation, with participants being less likely to judge the action as acceptable when personal force was involved ($p < .001$). Again, Japanese participants showed significantly lower acceptability of the action than participants from other regions (both $p < .001$), whereas intent and perspective were not significant predictors.

Additional likelihood-ratio tests examined whether the effects of Intent, Personal Force, and Perspective differed across regions. None of these interaction terms significantly improved model fit for moral choice (Q1: Intent × Region, $\Delta\chi^2(2) = 1.48$, $p = .478$; Personal Force × Region, $\Delta\chi^2(2) = 0.31$, $p = .857$; Perspective × Region, $\Delta\chi^2(2) = 0.62$, $p = .735$) or moral evaluation (Q2: Intent × Region, $\Delta\chi^2(2) = 3.76$, $p = .153$; Personal Force × Region, $\Delta\chi^2(2) = 1.42$, $p = .492$; Perspective × Region, $\Delta\chi^2(2) = 2.07$, $p = .355$).

Joint Modelling Across Judgement Types: To directly examine whether the effects of the key predictors differed across judgement domains, we conducted a generalised linear mixed-effects model (GLMM) in which responses from moral evaluation (Q2) and intentionality attribution (Q3) were jointly modelled in long format. Response Type (moral vs. intentionality) was included as a factor, along with its interactions with Intent (means vs. side effect), Personal Force, Perspective, and Region. Because each participant contributed responses to both judgement types, a random intercept for participants was included to account for repeated observations. However, the model yielded a singular fit, with the random-intercept variance estimated as zero.

The analysis revealed significant interactions between Response Type and all key predictors except Personal Force (Intent, Perspective, and Region), suggesting that the observed predictor patterns differed between moral judgement and intentionality attribution.

In particular, the direction and magnitude of some effects varied across judgement types. These exploratory differences are compatible with the possibility that the same factors play different roles

depending on the judgement being made. However, because the analysis was not preregistered, the measures were not fully matched in format, and the mixed-effects model yielded a singular fit, the results do not establish distinct or independent cognitive mechanisms.

Taken together, the exploratory analyses suggest possible differences in predictor patterns across judgement domains, but they should be treated as hypothesis-generating and tested in a purpose-built, preregistered study.

Discussion

The preregistered analyses yielded three central findings. First, the predicted Region by Intent and Region by Perspective interactions were not detected, so H1 and H3 were not supported. Second, Personal Force did not predict intentionality attribution, so H2 was not supported. Third, the pooled model showed main effects of intent and perspective and substantial differences in baseline attribution rates across regions. The simulation-based sensitivity analyses indicated that the present design had approximately 80% power only for relatively large interaction effects under the assumed directional patterns. The null interactions are therefore interpreted as inconclusive with respect to smaller or moderate cultural moderation effects rather than as evidence of cultural equivalence.

Hypotheses and Their Evaluation

Regarding our preregistered hypotheses, none of the three were consistent with our results.

H1 (Intent: the means/side-effect distinction): First, Intent was a significant predictor in the main-effects model, with harmful means generally judged as more intentional than harmful side effects. This pattern seems consistent with earlier literature, in which morally worse actions (presumably, in this case, the harm-as-means condition) tend to be judged as more intentional (Knobe 2003, 2006; Knobe and Mendlow 2004). However, the Region × Intent interaction was not statistically significant, and the analysis therefore did not support the cultural moderation predicted in H1. This differs from the cultural variation in the means/side-effect effect on moral permissibility reported by Bago et al. (2022), although the present null interaction does not establish cross-cultural equivalence.

H2 (Personal Force): Since earlier studies on personal force in trolley-type dilemmas suggested that moral and emotional processing differ between these actions (Bago et al. 2022; Greene et al. 2009), we expected that actions involving personal force might also be judged as more intentional. However, the preregistered analysis did not detect the predicted positive effect of Personal Force on intentionality attribution. Direct physical actions were not significantly more likely to be judged intentional than indirect actions. In secondary post hoc analyses, however, Personal Force significantly predicted moral choice (Q1) and moral evaluation (Q2).

H3 (Perspective)

Perspective significantly predicted intentionality attribution overall and in the separate regional models, but the Region × Perspective interactions were not significant. Thus, the study did not detect the predicted cultural moderation. The observed coefficient patterns are compatible with some similarity across the three samples, but the nonsignificant interactions do not demonstrate that the perspective effect is culturally invariant.

Secondary Exploratory Comparisons Between Judgement Types

The preregistered hypotheses concerned intentionality attribution. To describe how the observed Q3 pattern related to Q1 and Q2, we conducted secondary analyses comparing predictors across response types. These analyses suggest possible differences between intentionality attribution and moral judgement, but they were not preregistered and should be treated as hypothesis-generating.

In the observed data, intent and perspective predicted intentionality attribution, whereas personal force predicted moral choice and acceptability. We refer to this as a descriptive contrast in predictor patterns rather than as evidence of separate cognitive mechanisms. Baseline regional patterns also differed across response types.

The joint model provided an exploratory test of whether coefficients differed across response types. Several Response Type interactions were significant, but the model was not preregistered and produced a singular random-effects fit. Accordingly, it provides provisional evidence that should be tested in a purpose-built, adequately powered study rather than strong confirmation of a dissociation.

Revisiting Prior Findings in Moral Dilemmas

Although personal force and region predicted moral choice and moral evaluation, intent and perspective did not reliably predict those outcomes. This descriptive pattern suggests that the predictors of intentionality attribution and moral judgement may only partially overlap. Because these comparisons were post hoc, question order was fixed, and the response formats were not fully matched, the present study cannot determine whether the differences reflect distinct mechanisms, measurement differences, or order-related influences.

Regional Baseline Differences and Possible Linguistic Contributions

The observed similarity in some coefficient patterns across the three samples is compatible with the possibility of some shared features of intentionality attribution, but the nonsignificant interactions do not establish cross-cultural invariance. The large baseline difference for Chinese participants may reflect cultural, linguistic, sampling, or measurement factors. One possible linguistic explanation is that the Chinese term used for 'intentionally' (gu yi, 故意) does not fully overlap with the English term, as cross-linguistic work has shown that lexical counterparts can differ in evaluative meaning (Mizumoto 2018). This explanation remains speculative because lexical alternatives were not experimentally compared in the present study.

Limitations and Future Directions

Several limitations should be noted. First, the simulation-based sensitivity estimates were conditional on specific hypothesis-consistent interaction shapes and on nuisance parameters estimated from the observed data. Under those assumptions, the study achieved approximately 80% power only for relatively large Region × Intent and Region × Perspective interactions. Smaller or moderate culturally moderated effects may therefore have gone undetected, and the nonsignificant interactions cannot establish equivalence or the absence of cultural moderation.

Second, comparisons across Q1, Q2, and Q3, including the joint model, were not preregistered. The questions were presented in a fixed order, the response formats were not fully matched, and the mixed-effects model yielded a singular fit. These analyses should therefore be regarded as hypothesis-generating and require direct preregistered replication using matched response formats and randomised or experimentally manipulated question order.

Third, we used only one lexical item for intentionality for each language/culture, while, as noted above, the lexical meanings of the counterparts for intentionally may have affected participants' baseline attribution rates. So, future studies could address this by including conditions using gu yi and you yi (for Chinese), itoteki-ni and wazato (for Japanese), as well as additional languages for both cross-linguistic and intra-linguistic comparisons.

Fourth, we examined intentionality attribution only to the action, not to the omission (inaction). However, intentional omission is equally possible, and we would like to know how the patterns differ from those in this study.

Fifth, our study focused solely on harmful outcomes because of our interest in moral dilemma situations. Intentionality attributions might exhibit different patterns when the dilemma is not negative, where either choice produces a good outcome, or each outcome is morally neutral. Testing positive and neutral dilemma cases may reveal whether the patterns observed here reflect general cognitive tendencies or are specific to judgements about moral dilemmas.

Finally, although our participants were drawn from Japan, China, and the United States, these countries do not represent the full range of cultural diversity. Findings from these three groups may not generalise to other societies with different moral systems, languages, or models of agency. Future research should include participants from a wider set of cultures to build a more global picture of how intentional action is understood.

Conclusions

The present study provides a preregistered test of intentionality attribution in moral dilemmas across samples from the United States, Japan, and China. None of the three preregistered hypotheses was supported: cultural moderation of the Intent and Perspective effects was not detected, and Personal

Force did not significantly predict intentionality attribution. The pooled analyses nevertheless showed main effects of Intent and Perspective and substantial regional differences in baseline attribution rates.

These findings describe some similarity in the observed coefficient patterns across the present samples, but they do not establish cultural invariance. The simulation-based sensitivity analyses indicated that the design was primarily capable of detecting relatively large interactions under the assumed directional patterns. Secondary analyses suggested that predictors may operate differently across moral judgement and intentionality attribution. However, because these analyses were post hoc, the outcomes were formatted differently, and the mixed-effects model yielded a singular fit, the findings are best understood as directions for future confirmatory research rather than as evidence for distinct cognitive mechanisms.

The present data therefore motivate two focused next steps: preregistered studies designed to detect smaller interaction effects, and designs that randomise or manipulate question order while matching response formats across judgement types. Such studies can determine whether the patterns observed here are robust and whether they reflect cognitive, cultural, linguistic, or procedural factors.

Differences in lexical meaning remain one plausible explanation for the regional baseline patterns, but this account was not tested directly and should be evaluated by comparing alternative intentionality terms within languages.

In summary, the study identifies descriptive patterns within the present samples and clear targets for better-powered and more tightly controlled follow-up research.

Statements and declarations

Ethics declaration

The study was designed, coordinated, administered, and managed by researchers at the Japan Advanced Institute of Science and Technology (JAIST) in Japan. Participants in the three target samples took part remotely through online recruitment platforms. The study was conducted in accordance with the Declaration of Helsinki and applicable institutional guidelines. The study protocol was approved by the Knowledge Science Ethics Council at JAIST in Japan (Approval No. KSEC-G2024120604). Ethical approval was therefore obtained from the local ethics review body of the institution responsible for the research in Japan. The approved protocol explicitly included online recruitment for the three target samples associated with China, Japan, and the United States. Separate ethical approval was not obtained in China or the United States because recruitment for the relevant target samples and participation in the study were conducted entirely online, no local investigators, collaborating research institutions, or physical research sites were involved, and all study administration and data management were conducted by the research team at JAIST in Japan. All participants provided informed consent before participation. No personally identifiable information (PII) was collected.

Data availability

The study preregistration, study materials, anonymised data, analysis code, and simulation-based sensitivity analysis materials are publicly available on the Open Science Framework (Han and Mizumoto 2026a, 2026b).

Competing interests

The authors declare no competing interests.

Funding

This work was supported by JSPS KAKENHI Grant-in-Aid for Scientific Research (A) (Grant Number JP25H00459).

Author contributions

Z.H. conceived the study, developed the materials, conducted the analyses, and drafted the manuscript. M.M. supervised the study, contributed to the methodology, and critically revised the manuscript. Both authors approved the final manuscript.

Acknowledgements

Not applicable.

References

1. Awad E, Dsouza S, Shariff A, Rahwan I, Bonnefon J (2020) Universals and variations in moral decisions made in 42 countries by 70,000 participants. Proc Natl Acad Sci U S A 117(5):2332–2337. https://doi.org/10.1073/pnas.1911517117
2. Bago B, Kovacs M, Protzko J, Nagy T, Lange F, Wyszynski P et al (2022) Situational factors shape moral judgements in the trolley dilemma in Eastern, Southern and Western countries in a culturally diverse sample. Nat Hum Behav 6(7):880–895. https://doi.org/10.1038/s41562-022-01319-5
3. Choi I, Nisbett RE, Norenzayan A (1999) Causal attribution across cultures: Variation and universality. Psychol Bull 125(1):47–63. https://doi.org/10.1037/0033-2909.125.1.47
4. Gold N, Colman AM, Pulford BD (2014) Cultural differences in responses to real-life and hypothetical trolley problems. Judgm Decis Mak 9(1):65–76. https://doi.org/10.1017/S193029750000499X
5. Greene JD, Cushman FA, Stewart LE, Lowenberg K, Nystrom LE, Cohen JD (2009) Pushing moral buttons: The interaction between personal force and intention in moral judgment. Cognition 111(3):364–371. https://doi.org/10.1016/j.cognition.2009.02.001
6. Han Z, Mizumoto M (2026a) Intentionality attribution in trolley-type moral dilemmas across the United States, Japan, and China: preregistration. Open Science Framework. https://osf.io/an4se. Accessed 25 Jul 2026

7. Han Z, Mizumoto M (2026b) Intentionality attribution in trolley-type moral dilemmas across the United States, Japan, and China: study materials. Open Science Framework. https://osf.io/pehbt. Accessed 25 Jul 2026
8. Jiang Q, Zhuo L, Wang Q, Lin W (2022) The neural basis of moral judgement for self and for others: Evidence from event-related potentials. Front Hum Neurosci 16:919499. https://doi.org/10.3389/fnhum.2022.919499
9. Jones EE, Nisbett RE (1972) The actor and the observer: Divergent perceptions of the causes of behavior. In: Jones EE, Kanouse DE, Kelley HH, Nisbett RE, Valins S, Weiner B (eds) Attribution: Perceiving the causes of behavior. General Learning, Morristown, NJ, pp 79–94
10. Knobe J (2003) Intentional action and side effects in ordinary language. Analysis 63(3):190–194. https://doi.org/10.1093/analys/63.3.190
11. Knobe J (2006) The concept of intentional action: A case study in the uses of folk psychology. Philos Stud 130(2):203–231. https://doi.org/10.1007/s11098-004-4510-0
12. Knobe J, Burra A (2006) The folk concepts of intention and intentional action: A cross-cultural study. J Cogn Cult 6(1–2):113–132. https://doi.org/10.1163/156853706776931222
13. Knobe J, Mendlow GS (2004) The good, the bad and the blameworthy: Understanding the role of evaluative reasoning in folk psychology. J Theor Philos Psychol 24(2):252–258. https://doi.org/10.1037/h0091246
14. Leslie AM, Knobe J, Cohen A (2006) Acting intentionally and the side-effect effect. Psychol Sci 17(5):421–427. https://doi.org/10.1111/j.1467-9280.2006.01722.x
15. Malle BF (2006a) The actor-observer asymmetry in attribution: A (surprising) meta-analysis. Psychol Bull 132(6):895–919. https://doi.org/10.1037/0033-2909.132.6.895
16. Malle BF (2006b) Intentionality, morality, and their relationship in human judgment. J Cogn Cult 6(1–2):87–112. https://doi.org/10.1163/156853706776931358
17. Malle BF, Knobe J (1997) The folk concept of intentionality. J Exp Soc Psychol 33:101–121. https://doi.org/10.1006/jesp.1996.1314
18. Masuda T, Nisbett RE (2001) Attending holistically versus analytically: Comparing the context sensitivity of Japanese and Americans. J Pers Soc Psychol 81(5):922–934. https://doi.org/10.1037/0022-3514.81.5.922
19. Miller RM, Hannikainen IA, Cushman FA (2014) Bad actions or bad outcomes? Differentiating affective contributions to the moral condemnation of harm. Emotion 14(3):573–587. https://doi.org/10.1037/a0035361
20. Miyamoto Y, Kitayama S (2002) Cultural variation in correspondence bias: The critical role of attitude diagnosticity of socially constrained behavior. J Pers Soc Psychol 83(5):1239–1248. https://doi.org/10.1037/0022-3514.83.5.1239
21. Mizumoto M (2018) A simple linguistic approach to the Knobe effect, or the Knobe effect without any vignette. Philos Stud 175(7):1613–1630. https://doi.org/10.1007/s11098-017-0926-1

22. Morris MW, Peng K (1994) Culture and cause: American and Chinese attributions for social and physical events. J Pers Soc Psychol 67(6):949–971. https://doi.org/10.1037/0022-3514.67.6.949
23. Muthukrishna M, Bell AV, Henrich J, Curtin CM, Gedranovich A, McInerney J et al (2020) Beyond Western, educated, industrial, rich, and democratic (WEIRD) psychology: Measuring and mapping scales of cultural and psychological distance. Psychol Sci 31(6):678–701. https://doi.org/10.1177/0956797620916782
24. Nadelhoffer T, Feltz A (2008) The actor-observer bias and moral intuitions: Adding fuel to Sinnott-Armstrong's fire. Neuroethics 1(2):133–144. https://doi.org/10.1007/s12152-008-9015-7
25. 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. Curr Psychol 43(6):5243–5253. https://doi.org/10.1007/s12144-023-04662-6
26. Robbins E, Shepard J, Rochat P (2017) Variations in judgments of intentional action and moral evaluation across eight cultures. Cognition 164:22–37. https://doi.org/10.1016/j.cognition.2017.02.012
27. Sauer H (2014) It's the Knobe effect, stupid! How (and how not) to explain the side-effect effect. Rev Philos Psychol 5:485–503. https://doi.org/10.1007/s13164-014-0189-0

Figures

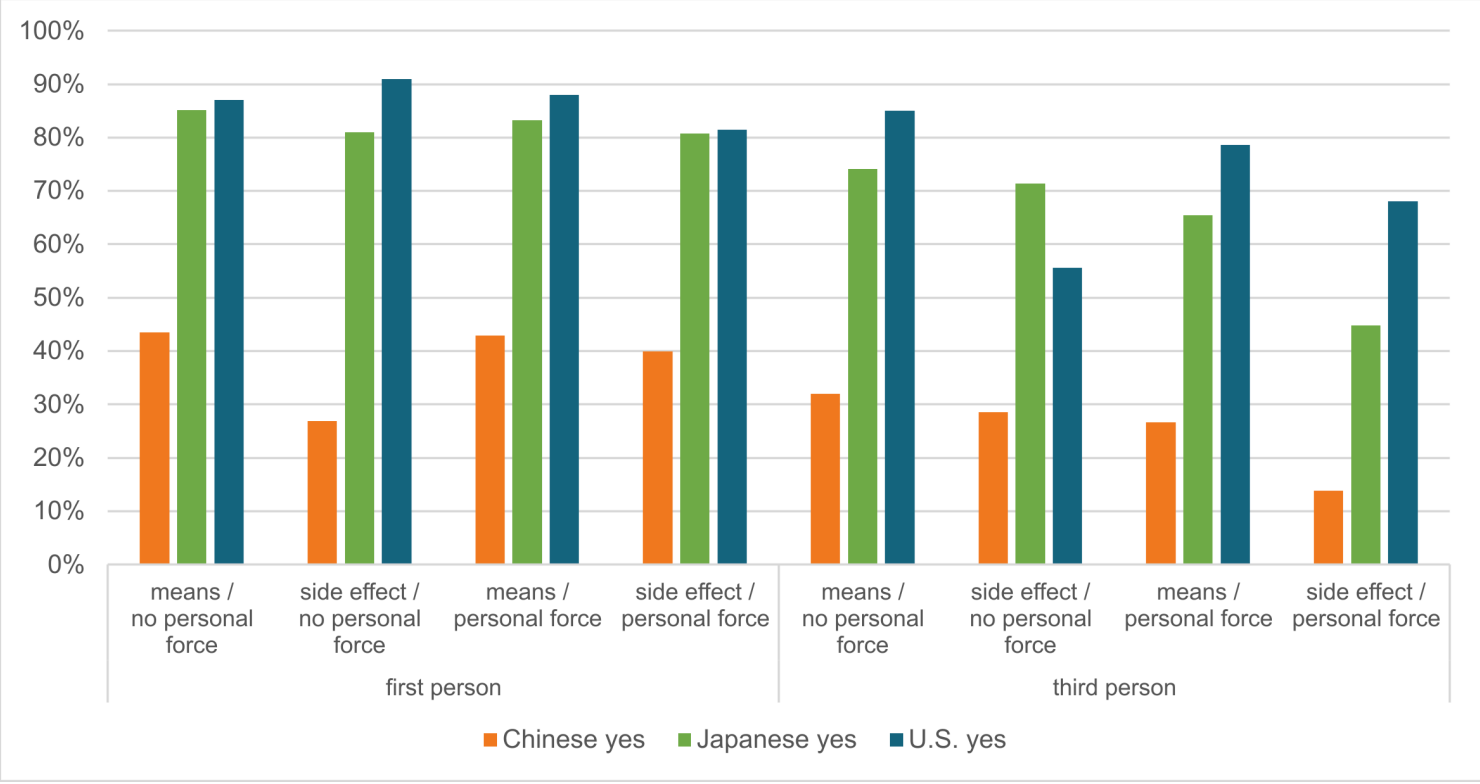


Figure 1

Intentionality attribution rates by experimental condition and sample. Bars show the percentage of “yes” responses to Q3 (intentionality attribution) for Chinese, Japanese, and U.S. participants. Conditions vary by intent (means vs. side effect), personal force (push vs. switch), and perspective (first person vs. third person); no error bars are shown.