# Paradigm Evolution in Social Dilemma Research and Its Theoretical Value: The Case of Moral Dilemma Decision-Making

**Authors:** Liu Chuanjun, Liao Jiangqun, Liao Jiangqun

**Date:** 2019-11-15T00:00:00+00:00

## Abstract

Social dilemmas constitute a common research theme across multiple disciplines including ethics, sociology, psychology, and economics, with moral dilemma decision-making as a representative exemplar. Based on a comprehensive analysis of the evolutionary trajectory of three empirical research paradigms in moral decision-making studies (the classic dilemma approach, the process dissociation approach, and the CNI model approach), this paper proposes the CAN dimension synthesis method, which resolves the theoretical and methodological limitations inherent in the aforementioned research paradigms. The development of these four methodological approaches enables moral dilemma research to transcend the confines of conflicting dilemma situations, thereby expanding the scope of moral scenario investigation, facilitating empirical validation of existing moral theories and resolution of scholarly controversies, and offering methodological references for research topics with potential conflicts in related disciplines.

## Full Text

## Preamble

The Evolution of Paradigms and Their Theoretical Values in Social Dilemma Research: The Case of Moral Dilemma Decision-Making Liu Chuanjun and Liao Jiangqun* (Department of Psychology, School of Social Sciences, Tsinghua University, Beijing 100084)

*This research was supported by the National Social Science Fund Project (18BSH114) and Tsinghua University's Independent Research Program (2017THZWYY11).*

Corresponding author: Liao Jiangqun, E-mail: liaojq@tsinghua.edu.cn.

## Abstract

Social dilemma is a common research topic across multiple disciplines including ethics, sociology, psychology, and economics, with moral dilemma decision-making as its representative paradigm. Based on a comprehensive analysis of the evolution of three empirical research paradigms in moral decision-making research (the classic dilemma method, process dissociation method, and CNI model method), this paper proposes the CAN dimensional synthesis method, which addresses the theoretical and methodological limitations of the aforementioned research paradigms. The development of these four methods has enabled moral dilemma research to move beyond the confines of conflicting dilemma situations, thereby expanding the scope of moral situation research, facilitating empirical testing of existing moral theories and resolution of research controversies, and providing methodological references for potentially conflicting research topics in related disciplines.

**Keywords:** social dilemma; moral dilemma; rule conflict; paradigm evolution; empirical methods

With the advent of autonomous vehicles, social dilemma issues have once again attracted widespread attention. Unlike previous hypothetical scenarios that seemed unlikely to occur in reality, the arrival of autonomous vehicles has brought dilemma problems into tangible existence. [1] When an autonomous vehicle encounters an emergency situation during operation—if it continues on its original route, it will kill five or more pedestrians, but if it changes direction to hit the curb, it will sacrifice the vehicle and its passengers—how should the vehicle respond in such emergencies? The programming behind autonomous vehicles essentially reflects people' s decision-making tendencies. Recent studies have found that while people tend to sacrifice the vehicle and its passengers, they are unwilling to ride in or purchase such vehicles. [2][3]

Similar to autonomous vehicles, social dilemmas first emerged in the moral domain as moral dilemma scenarios and later expanded to various aspects of social life. The rise of moral dilemma research has its background in both positivism and folk morality. Before the formation of the moral dilemma research paradigm, moral research had always relied on philosophical speculation. This approach has at least three limitations: First, philosophers from different schools hold different fundamental positions, leading to inherent differences in their basic views on morality and making it difficult to reach consensus through debate. Second, morality under philosophical speculation may not align with morality in the eyes of ordinary people, yet morality from a folk perspective holds greater practical and social significance. Third, these moral perspectives remain at the level of philosophical speculation and are not empirically testable, preventing direct objective verification. Against this backdrop, moral research urgently needed to develop an intuitive, testable, and easily understandable research paradigm for ordinary people, giving rise to thought experiments in moral dilemmas.

In 1967, Foot proposed the famous trolley problem, marking the formation

of the moral dilemma research paradigm. [1] Approximately ten years later, Thomson proposed another classic moral dilemma—the footbridge paradigm. [2] In the trolley problem, decision-makers must choose whether to flip a switch to divert a trolley, thereby deciding whether to sacrifice one innocent person to save five. In the footbridge paradigm, decision-makers must choose whether to push a large stranger off a bridge to save five workers. Endorsing the proposal to sacrifice one (or a few) to save five (or many) is considered a utilitarian moral judgment because it is based on outcome maximization, aligning with the utilitarian principle of maximizing human welfare. Conversely, rejecting the proposal is considered a deontological moral judgment because it is believed to be based on whether the action itself conforms to moral norms, consistent with deontological principles regarding the normative validity of actions.

The implementation method in the trolley problem is relatively indirect—the innocent person' s death is caused directly by the trolley rather than by the decision-maker' s own hand—hence it is called an impersonal dilemma. In contrast, in the footbridge paradigm, the innocent person' s death results from the decision-maker directly pushing them, thus it is called a personal dilemma. These two types of dilemmas have become classics in subsequent moral research, with numerous studies examining people' s decision preferences and their antecedents and consequences in these dilemmas and their variants. [3][4][5][6] For instance, some researchers have found that men are more utilitarian in personal moral dilemmas, but no gender differences exist in impersonal dilemmas. [7] Linguistic framing may also play a role—for example, when the action proposal is described as "save lives" versus "do not kill," people' s decision tendencies differ. [8] For details, see the review by Xu Tongjie et al. [9] Moral dilemmas have been widely applied, and in China' s moral education, the moral dilemma story method is often used for classroom discussions to enhance students' moral cognitive development. [10][11]

As research has progressed, numerous moral dilemmas and their variants have been developed, with researchers employing diverse situational materials and lacking standardized research methods, which has prevented many meta-analyses from being conducted due to the absence of methodological common ground. [1] Additionally, studies have pointed out that the word count, expression methods, and question formats of situational materials all influence moral judgments and require standardization in research. [11] More importantly, behind the use of materials themselves, there exist differences in the empirical methods of these moral dilemma studies. Clarifying the methodological paradigms in moral dilemma research is of great significance for both moral dilemma research and similar social dilemma studies. This paper will provide a comprehensive analysis of the currently mainstream classic dilemma method, process dissociation method, and CNI (Consequence, Norm, and generalized Inaction preference) model method. Based on this analysis, we propose the CAN (Consequence sensitivity, generalized Action preference, and Norm sensitivity) dimensional synthesis method and discuss its underlying methodological value.

## II. Three Existing Dilemma Research Paradigms

### (1) Classic Dilemma Method

The classic dilemma method primarily uses situational stories such as the trolley problem and footbridge paradigm and their variants, asking people to evaluate the moral legitimacy of action proposals and their willingness to act in these hypothetical scenarios, thereby exploring people' s moral tendencies. In these hypothetical situations, one innocent person (or a few) must be sacrificed to save five (or many). Decision-makers are asked whether such actions are morally acceptable and whether they would be willing to execute the proposed action. For example, in the footbridge paradigm, decision-makers first read: "A runaway trolley is rushing toward five workers on the tracks. If the trolley continues, they will be killed. You are on a footbridge above the tracks, positioned between the speeding trolley and the five workers. Next to you on the footbridge stands a large stranger. The only way to save the five workers' lives is to push this stranger off the bridge onto the tracks below. His large body will stop the trolley. If you do this, the stranger will die, but the five workers will be saved." Decision-makers are then asked, "Do you think the proposal to 'push the stranger to save the five workers' is morally acceptable?" or "Would you push the stranger to save the five workers?"

The dual-process theory of morality was developed using the classic moral dilemma method. Previous moral theories were primarily based on rational cognition, represented by Piaget [3] and Kohlberg [4], who argued that rational reasoning forms the basis of moral judgment. However, Greene discovered that variations in emotional engagement influence moral judgment and initially discussed the neural mechanisms of moral judgment using fMRI methods. In Greene' s early research, he collaborated extensively with moral philosopher Haidt, emphasizing the moral utility of emotion and demonstrating through experiments that, beyond rational deliberation, emotion plays an important role in influencing moral judgment, particularly in predicting individuals' deontological tendencies. [5][6] Later, Greene also discovered that rational thinking plays a similarly important role in moral judgment, particularly in predicting utilitarian tendencies. [1][2] Consequently, Greene proposed the dual-process theory of morality, [3] which has had a profound impact on moral research, with numerous subsequent studies conducted within this theoretical framework. [4][5][6] For reviews of this theory, see the analysis by Yu Feng et al. [7]

The classic dilemma method and the dual-process theory of morality developed from it have revealed many counterintuitive, divergent, or even contradictory results in subsequent research. First, regarding utilitarian response tendencies, studies have found positive correlations between utilitarian responses and decision-makers' psychopathy, Machiavellianism, and sense of meaninglessness in life, [8] with the positive correlation between psychopathy and utilitarian responses being repeatedly confirmed in follow-up studies. [9][10][11] The weighing of pros and cons of behavioral outcomes has long been considered a manifesta-

tion of utilitarian rationality, and dual-process theory posits that rational processing drives utilitarian responses. However, this is positively correlated with psychopathy, implying that stronger psychopathy or psychopathic individuals are more rational and concerned with maximizing human welfare compared to normal people, or that more rational people concerned with maximizing human welfare may have stronger psychopathic tendencies—both of which are theoretically and commonsensically absurd.

Second, regarding deontological response tendencies, the positive correlation between disgust and deontological tendencies has been examined in many studies, [12][13] with most arguing that disgust stimuli evoke decision-makers' feelings of aversion toward behaviors that violate moral norms, particularly showing stronger aversion toward violations of purity, thereby driving stricter moral standards and more deontological moral judgments. [1][2][3] On the other hand, positive emotions can counteract some disgust feelings, thereby reducing deontological responses and increasing utilitarian responses. [4] However, some studies have noted that two types of positive emotions—mirth and elevation—have completely different effects, with the former decreasing deontological tendencies while the latter has the opposite effect. [5] Therefore, the relationship between disgust emotion and deontological tendencies may be more complex and difficult to explain reasonably within the methodological framework of the classic dilemma method.

Third, is there a strict correspondence between intuitive/rational processing and deontological/utilitarian responses? From the perspective of mental model theory, intuitive utilitarianism is possible, [6] and recent studies using a two-response paradigm have found that individuals may be utilitarian at the intuitive level. [7][8] Individuals' utilitarian judgment outcomes remain largely unchanged between intuitive processing states and rational processing states when making judgments a second time. Moreover, the Cognitive Reflection Test does not show a significant correlation with utilitarian responses. [9] Therefore, the correspondence between intuitive/rational processing and deontological/utilitarian responses requires re-examination. Indeed, some critics argue that so-called utilitarian moral judgments within the dilemma framework cannot reflect people' s concern for maximizing outcomes and further question the validity of the classic dilemma method in revealing people' s moral choices. [10]

The greatest limitation of the classic dilemma method is that it strictly opposes deontological and utilitarian principles, operationally equating decision-makers' acceptance of utilitarian proposals with being driven by utilitarian principles, and equating rejection of utilitarian proposals with being driven by deontological principles. This operationalization creates several limitations. First, it treats utilitarian and deontological tendencies as incompatible—the stronger the utilitarian tendency, the weaker the deontological tendency, and vice versa. This contradicts both common sense and recent research findings. Common sense suggests that when facing a moral choice, people can simultaneously consider both the moral norms behind an action and the moral consequences it will

produce. Recent neuroscientific evidence indicates that the anterior cingulate cortex, insula, and superior temporal gyrus are associated with emotional evaluation, while the temporoparietal junction and dorsomedial prefrontal cortex are associated with utilitarian evaluation, with overall moral value judgment represented in the anterior portion of the ventromedial prefrontal cortex. Crucially, the response patterns and functional interactions among these three sets of regions suggest that emotional and utilitarian evaluations are computed independently and in parallel before being integrated into an overall moral value judgment in the ventromedial prefrontal cortex. [1] This implies that utilitarian and deontological response tendencies can be completely independent and parallel.

The second limitation of the classic dilemma method is its inability to quantify the degree differences in decision-makers' utilitarian and deontological response tendencies. When these two tendencies differ substantially, decision-makers can make judgments immediately; when they differ only slightly, decision-makers require more time but may still reach the same judgment. Particularly in binary choice situations where decisions must either accept or reject utilitarian proposals, these degree differences in response tendencies cannot be captured. [2]

The third limitation of the classic dilemma method lies in its explanatory ambiguity. [3] By strictly opposing utilitarian and deontological principles, when utilitarian responses increase (for example, when the probability of choosing to flip the switch in trolley-like dilemmas increases), this can be interpreted either as a strengthening of utilitarian tendencies or as a weakening of deontological tendencies, creating interpretational ambiguity. Furthermore, decision-makers' endorsement of utilitarian proposals may result either from utilitarian principle-driven motivation or merely from a general preference for accepting proposals (for instance, in trolley problems, decision-makers might accept the proposal to flip the switch regardless of moral norms or outcome favorability, simply due to a general acceptance preference). Similarly, rejecting utilitarian proposals may result either from deontological principle-driven motivation or from a general rejection preference.

To address the first two limitations, researchers proposed the Process Dissociation (PD) method, while to address all the aforementioned limitations, researchers developed the CNI model method and the CAN dimensional synthesis method proposed in this paper.

### (2) Process Dissociation Method

The process dissociation method was developed by Conway and Gawronski from the process dissociation procedure (PD) used in memory research. [4] Jacoby first used the PD method to separate recollection components from familiarity-based guessing components in memory performance, and this method is not content-specific and can be applied in many domains. [5] Conway and Gawronski were the first to introduce it into moral psychology research to distinguish the

degree of influence from utilitarian and deontological principles. [6]

The process dissociation method essentially manipulates the benefit-cost ratio of action outcomes. When benefits outweigh costs, it creates an incongruent situation with the harmful proposal—meaning that although the outcome is favorable, the action itself violates moral norms, with utilitarian principles requiring acceptance and deontological principles requiring rejection. When costs outweigh benefits, it creates a congruent situation where the outcome is harmful and the action violates moral norms, with both utilitarian and deontological principles requiring rejection. The principle is illustrated in Figure 1 [Figure 1: see original paper]. Incongruent situations are equivalent to classic dilemma situations such as the aforementioned trolley problem and footbridge paradigm. Congruent situations adjust the outcome of the same scenario so that costs outweigh benefits —for example, in the trolley problem, if the runaway trolley would kill only one worker on its original track but killing five workers if the switch is flipped to divert it to another track, then regardless of whether utilitarian or deontological principles are driving the decision, this harm would be unacceptable. However, if neither principle is activated, people might consider this harm acceptable.

In terms of calculation principles, the process dissociation method uses multiple sets of scenarios, with each set forming two variants—congruent and incongruent —by adjusting the benefit-cost ratio of action outcomes, presented in the form of decision processing trees. The method then calculates the degree of utilitarian and deontological tendencies by computing decision-makers' probabilities of accepting or rejecting harmful proposals in congruent and incongruent situations. Since there are multiple scenario stories under both congruent and incongruent conditions, researchers can calculate decision-makers' response probabilities for acceptance and rejection separately under the two conditions. These response probabilities are then used to derive the degree of influence from utilitarian and deontological principles. The specific calculations are as follows:

p(reject harm|congruent) = U + (1 − U) × D p(accept harm|congruent) = (1 − U) × (1 − D) p(reject harm|incongruent) = (1 − U) × D p(accept harm|incongruent) = U + (1 − U) × (1 − D)

The left side of equations (1)~(4) represents the probabilities of decision-makers accepting or rejecting action proposals in congruent and incongruent situations, which can be calculated directly from decision outcomes. For example, if there are 8 scenario stories in an incongruent situation and the decision-maker chooses to accept in 6 of these stories, then p(accept harm|incongruent) = 0.75 and p(reject harm|incongruent) = 0.25. Similarly, acceptance and rejection probabilities can be calculated for congruent situations. The values of U and D can then be derived from equations (1)~(4). For instance, by subtracting equation (2) from equation (4), or equation (3) from equation (1), we obtain:

U = p(accept harm|incongruent) − p(accept harm|congruent) U = p(reject harm|congruent) − p(reject harm|incongruent)

After calculating the value of U, the value of D can be computed by combining

it with any one of equations (1)~(4).

Compared to the classic dilemma method, the process dissociation method represents a theoretical and methodological advancement, expanding from the single-type situations of the classic dilemma method to dual-type situations. The primary theoretical significance of this expansion includes:

First, the process dissociation method breaks the one-to-one correspondence between behavioral responses and their underlying moral principles that exists in the classic dilemma method. In the classic dilemma method, utilitarian responses (i.e., accepting utilitarian proposals) correspond to being driven by utilitarian principles, while deontological responses (i.e., rejecting utilitarian proposals) correspond to being driven by deontological principles. The process dissociation method breaks this correspondence, allowing both utilitarian and deontological tendencies behind any choice to be measured rather than being mutually exclusive.

Second, the process dissociation method introduces congruent dilemma situations as a reference, thereby separating out situations where neither utilitarian nor deontological principles drive behavioral responses—the situation represented by the bottom row in Figure 1. This situation had been confounded with utilitarian tendency manifestations in the classic dilemma method.

Conway and Gawronski conducted three empirical studies while developing this method, with results further supporting the dual-process theory of morality. They found that deontological tendencies are indeed rooted in emotional responses to harmful actions: individual differences in empathic concern and perspective-taking correlate with the D parameter reflecting deontological tendencies but not with the U parameter reflecting utilitarian tendencies; enhancing empathic concern affects only the D parameter, not the U parameter; individual differences in need for cognition correlate with the U parameter but not the D parameter, and cognitive load manipulations selectively affect only the U parameter without influencing the D parameter. [1]

The process dissociation method provides additional insights for many inconsistent conclusions derived from classic dilemma method research. For example, some studies using the classic dilemma method have suggested that men are more utilitarian than women. Within the classic dilemma framework, this could mean either stronger utilitarian tendencies or weaker deontological tendencies. However, using process dissociation comprehensively reveals that women have moderately higher deontological tendencies than men ($d = 0.57$), while men have only slightly higher utilitarian tendencies than women ($d = 0.10$). This indicates that gender differences in moral dilemmas primarily result from differences in emotional responses to harm rather than differences in cognitive evaluation of outcomes. [2] Another example involves research suggesting that sacrificial judgments reflect antisociality rather than genuine utilitarianism. [3] The process dissociation method reveals that antisociality predicts decreases in deontological tendencies but cannot predict increases in utilitarian tendencies,

and that ordinary people' s sacrificial utilitarian judgments also reflect moral concern for reducing harm. Whether philosophers or ordinary people, sacrificial utilitarian judgments reflect genuine moral concern. [4]

Given that utilitarianism requires maximizing outcomes while deontology emphasizes that actions must conform to moral norms, research should manipulate differences in outcome benefits for the former and manipulate whether action proposals conform to moral norms for the latter. Therefore, at least four possible combinations emerge: (1) action proposals have benefits greater than costs but are prohibited by norms; (2) action proposals have costs greater than benefits and are prohibited by norms; (3) action proposals have benefits greater than costs and are prescribed by norms; (4) action proposals have costs greater than benefits but are prescribed by norms.

Although the process dissociation method advanced moral dilemma research theoretically and methodologically, its theoretical framework remains incomplete. Among the four possible combinations of norms (prohibited or prescribed) and outcomes (benefits greater than costs or costs greater than benefits), it covers only two combinations, lacking examination of the other two. This deficiency creates two limitations: First, it cannot separate decision-makers' general tendencies to reject or accept action proposals. In actual decision-making behavior, there exists a type of decision behavior that does not concern itself with the norms or outcomes behind the decision but merely expresses general acceptance or rejection of the proposal—a tendency that cannot be captured by the process dissociation method. Second, just as the classic dilemma method confounds general acceptance tendencies with utilitarian tendencies, the process dissociation method confounds general rejection tendencies with deontological tendencies. In the second row of the decision processing tree in Figure 1, decision-makers choose to reject action proposals in both congruent and incongruent situations, which could be driven either by deontological principles or by a general rejection of action proposals regardless of whether norms are appropriate or outcomes are favorable. To address this important limitation, Gawronski' s research team developed the CNI model method in 2017 using multinomial decision tree models from econometrics, which separates not only the degree to which decisions are driven by utilitarian (or consequentialist) principles and deontological (or normative) principles but also the degree of decision-makers' general rejection/acceptance tendencies irrespective of norms or outcomes.

### (3) CNI Model Method

Building upon the process dissociation method, the CNI model method operationalizes utilitarian tendencies as sensitivity to consequences and deontological tendencies as sensitivity to norms, while further separating general rejection/acceptance tendencies. Theoretically, it covers all four combinations of norms (prohibited or prescribed) and outcomes (benefits greater than costs or costs greater than benefits), as shown in Table 1. The multinomial decision tree model forms the foundation of the CNI model method and has wide ap-

plications in many areas of social psychology. [1] The CNI model method can simultaneously quantify decision-makers' sensitivity to consequences (C parameter), sensitivity to norms (N parameter), and general preference for inaction versus action irrespective of consequences and norms (I parameter). Therefore, this method is called the CNI model of moral decision-making.

**Table 1** Example scenarios used in the CNI model method [2]

| Norm Prohibited | Norm Prescribed |
|---|---|
| **Benefits > Costs (p1):** You are an ordinary traveler. In the train station ticket hall, a person needs to buy a ticket for a train that is about to depart, and this is the only train of the day. If you help him cut in line to buy the ticket and board the train, his group of five can travel smoothly, but it will affect another person' s travel time and prevent them from buying a ticket. You plan to help him cut in line. | **Benefits > Costs (p3):** You are an ordinary traveler. In the train station ticket hall, a person needs to buy a ticket for a train that is about to depart and is currently cutting in line. If you stop him from cutting in line, he will not be able to travel on time, and the chaos caused by the stopping process will also prevent the five people in line from buying tickets on time. You plan to stop him from cutting in line. |
| **Costs > Benefits (p2):** You are an ordinary traveler. In the train station ticket hall, a person needs to buy a ticket for a train that is about to depart and is currently cutting in line. If you stop him from cutting in line, he will not be able to travel on time, but you can ensure that approximately five people in line can successfully buy tickets for the only train of the day. You plan to stop him from cutting in line. | **Costs > Benefits (p4):** You are an ordinary traveler. In the train station ticket hall, a person needs to buy a ticket for a train that is about to depart, and this is the only train of the day. If you help him cut in line to buy the ticket, he can successfully purchase it and board the train, but this will cause the five people in line to be unable to buy tickets for the only train of the day and thus unable to travel normally. You plan to help him cut in line. |

The CNI model method assumes that decision-makers follow sequential processing rules when making moral decisions: they first consider whether the outcome of the action proposal is favorable, then consider whether the proposal conforms to norms, and if neither is considered, they will choose to generally accept or reject the action proposal. Therefore, using decision processing tree models can peel away these different tendencies layer by layer, as shown in Figure 2 [Figure 2: see original paper]. When driven by utilitarian principles, decision-makers accept proposals when outcomes are favorable and reject them when outcomes are unfavorable (first row of Figure 2). If not driven by utilitarian principles but

by deontological principles, decision-makers accept proposals when norms prescribe them and reject them when norms prohibit them (second row of Figure 2). If driven by neither utilitarian nor deontological principles, decision-makers will choose to generally reject (third row of Figure 2) or accept (fourth row of Figure 2) the action proposal.

**Figure 2.** Multinomial decision processing tree diagram of the CNI model, where C represents sensitivity to consequences, N represents sensitivity to norms, and I represents general preference for inaction (the sum of response probabilities for general preference for action and general preference for inaction equals 1, so only one I parameter is needed in the model) [1]. Liu Yuanyuan et al. have also introduced this model in Chinese. [2]

p(accept|norm prohibited, benefits>costs) = C + (1 − C) × (1 − N) × (1 − I) p(accept|norm prohibited, costs>benefits) = (1 − C) × (1 − N) × (1 − I) p(accept|norm prescribed, benefits>costs) = C + (1 − C) ×N + (1 − C) × (1 − N) × (1 − I) p(accept|norm prescribed, costs>benefits) = (1 − C) ×N + (1 − C) × (1 − N) × (1 − I)

After multiple scenario stories under the same type of situation, the sum of probabilities for accepting and rejecting equals 1. For example, in the situation where norms are prohibited and benefits outweigh costs, if there are 6 scenario stories and the decision-maker chooses to accept the action proposal in 2 of these stories, then the acceptance probability is 1/3 and the rejection probability is 2/3. Therefore, the probability equations for acceptance and rejection reactions under the same type of situation are statistically equivalent. Here, only the probability equations for decision-makers accepting proposals under the four combined situations are listed. To simplify expressions, we sequentially designate p(accept|norm prohibited, benefits>costs) as p1, p(accept|norm prohibited, costs>benefits) as p2, p(accept|norm prescribed, benefits>costs) as p3, and p(accept|norm prescribed, costs>benefits) as p4, with the same notation used below.

To minimize the difference between the empirically observed probabilities of decision-makers' acceptance/rejection responses in the four situations and the probabilities predicted by the model equations using specified parameter estimates, the multinomial decision tree model employs maximum likelihood estimation to achieve this goal. The accuracy of model fit to data can be tested using means of goodness-of-fit statistics. Poor model fit indicates large variation between empirically observed probabilities and model-estimated probabilities, violating the basic assumptions of the model. The degree of model fit variation can be expressed by calculating Cohen' s w, where 0.1 indicates small variation, 0.3 indicates medium variation, and 0.5 indicates large variation. [1] However, when sample sizes are very large, examination of model fit variation may be unnecessary, as increased sample size itself leads to increased model fit variation.

Using the decision tree tool provided by Gawronski, researchers can directly cal-

culate decision-makers' sensitivity to consequences (C parameter), sensitivity to norms (N parameter), and general preference for inaction/action (I parameter) (http://www.bertramgawronski.com/documents/CNI-Model_Materials.zip). Additionally, these three parameters can be compared between two groups of data or against a specific reference value. If the obtained C and N parameters are significantly greater than 0, this indicates that decision-makers have significant sensitivity to consequences and norms. If the I parameter is significantly greater than 0.5, this indicates a significant general preference for inaction; if significantly less than 0.5, it indicates a significant general preference for action.

As previously mentioned, to address the limitations of the classic dilemma and process dissociation methods, Gawronski' s team developed the CNI model method. [2][3] This method also promotes the resolution of controversial issues in moral dilemma research.

For example, previous research found that emotion plays an important role in moral judgment but could not determine whether emotion acts on norm sensitivity, consequence sensitivity, or merely on general response tendencies. Using the CNI model method, studies have shown that happiness reduces sensitivity to moral norms without affecting sensitivity to consequences or general preference for inaction, while sadness and anger have no significant effects on moral dilemma judgments. [1] Previous research indicated that testosterone enhances decision-makers' utilitarian response tendencies, but using the CNI model method, researchers found that exogenous steroid testosterone enhances decision-makers' tendency to reject action proposals in situations where norms are prohibited and benefits outweigh costs, a response caused by increased sensitivity to norms. Conversely, the pattern for endogenous testosterone at baseline measurement is exactly opposite—the higher the endogenous testosterone level, the lower the sensitivity to moral norms. These results indicate that the role of testosterone in moral judgment is more complex than previously reported. [2] These studies demonstrate the methodological value of formal modeling approaches like the CNI model for deeply exploring the determinants of moral judgment.

Although the CNI model method comprehensively covers all four combinations of norms (prohibited or prescribed) and outcomes (benefits greater than costs or costs greater than benefits) in its theoretical constructs, making breakthrough methodological contributions, this method still has several limitations:

First, the method cannot be applied to correlational or regression research designs. The underlying foundation of the CNI model method is the multinomial decision tree model, which is widely used to distinguish multiple influencing factors in antecedent variables. However, this method is based on group comparisons rather than individual comparisons, and the generated C, N, and I parameters are represented at the group level rather than the individual level. Therefore, it cannot be used for correlational or regression analyses.

Second, limited by the multinomial decision tree tool, it can only compare differences between two groups of parameters or against a reference value, but cannot conduct comparisons among multiple groups. Consequently, the expandability of its computational method is relatively restricted.

Third, the CNI model method cannot conduct simple effects analysis and cannot more deeply analyze which of the four structural situations the differences between cross-group data exist in. For example, Gawronski et al. (2017) found that under action framing questions compared to judgment framing questions, decision-makers' norm sensitivity decreases while general preference for inaction increases. However, the CNI model method cannot provide further answers regarding which of the four structural situations this framing effect exists in.

Finally, the most significant problem with the CNI model method is that its theoretical logic a priori assumes that decision-makers' decision-making process follows a specific sequential processing pattern: decision-makers first consider outcomes, then consider norms if outcomes are not considered, and only exhibit general rejection or acceptance tendencies when neither outcomes nor norms are considered. The construct logic of the CNI model method is layer-by-layer stripping rather than parallel construction. In reality, decision-makers could very well process decisions in parallel, simultaneously considering both the norms behind the decision and the potential outcomes of the decision, or they might first consider whether the decision conforms to norms and then consider outcomes, or perhaps first form a general behavioral attitude that is then modified by normative or consequentialist principles. Therefore, if the positions of parameters in Figure 2 are swapped, the resulting probabilistic model of the algorithm will be completely different, as detailed below.

## III. Newly Developed Paradigm–CAN Dimensional Synthesis Method

As previously mentioned, the CNI model method has made breakthrough theoretical and methodological contributions based on the original classic dilemma and process dissociation methods. However, the CNI model method' s theory and algorithm also have limitations. Particularly in theoretical constructs, it assumes that decision-makers first consider behavioral outcomes, then consider behavioral normativity, and finally consider general behavioral tendencies in a hierarchical sequence. This fundamental assumption in the construct has not yet been empirically tested.

If the sequential processing that the CNI model method presupposes decision-makers follow conforms to reality, then besides processing decisions according to the outcome→norm→general rejection/acceptance tendency sequence assumed by the CNI model method, decision-makers might also process decisions according to norm→outcome→general rejection/acceptance tendency, or general rejection/acceptance tendency→modified by norms→modified by outcomes, or general rejection/acceptance tendency→modified by out-

comes→modified by norms, among other sequential processing approaches. Taking the norm→outcome→general rejection/acceptance tendency sequence as an example, when decision-makers process decisions according to this sequence, the decision tree model is as shown in Figure 3 [Figure 3: see original paper]:

**Figure 3.** Decision processing tree model under norm→outcome→general rejection/acceptance tendency

Correspondingly, decision-makers' decision response probabilities under various situations are: p1 = (1 − N) × C + (1 − N) × (1 − C) × (1 − I) p2 = (1 − N) × (1 − C) × (1 − I) p3 = N + (1 − N) × C + (1 − N) × (1 − C) × (1 − I) p4 = N + (1 − N) × (1 − C) × (1 − I)

If, as assumed by the CNI model method, decision-makers'decision processing sequence has no effect on their decision tendency estimates, then equations (5)~(8) should be statistically equivalent to equations (9)~(12). However, simple conversion reveals that this scenario has minimal probability of holding statistically. For example, converting the N parameter in the probability equations of the Figure 2 model yields N = (− p1 − p2 + p3 + p4) ⁄ (2 − p1 + p2 − p3 + p4), whereas converting the N parameter in the probability equations of the Figure 3 model yields N = (− p1 − p2 + p3 + p4) ⁄ 2. If the positional relationships of parameters in the CNI model do not affect probability estimation results, these two N parameters should be equal. This would lead to the conversion p2 − p1 = p3 − p4, which has minimal probability of holding in empirical statistics. Gawronski et al. (2017) reported in footnote 7 that when the positions of C and N are swapped, the effect results of their study can be replicated, but results that were marginally significant become significant. Therefore, the authors did not conduct further analysis on this, nor did they explain why statistical significance increased. [1] However, this precisely demonstrates that the positions of C and N do affect final parameter estimation results, but this effect was not further examined because it happened to support the authors' hypotheses.

Furthermore, Gawronski et al. (2017) argued that general behavioral response tendencies can only exist at the lowest level of the model because their response probabilities sum to 1, representing mutually contradictory response tendencies. This a priori assumption may not hold either. Decision-makers could very well first possess general acceptance or rejection tendencies that are then modified by normative or consequentialist principles, changing their original response tendencies. In other words, the positional relationships among the three parameters C, N, and I in Figures 2 and 3 are logically uncertain and interchangeable.

From the above analysis, we can conclude that if decision-makers' decision patterns follow sequential processing forms, multiple decision-making modes should exist. The CNI model algorithm obviously targets only one of these modes, which empirically may indeed be the decision-making mode of most decision-makers. However, we cannot ignore the possibility of other decision-making modes existing, and decision-makers are indeed more likely to weigh norms and

outcomes simultaneously, which creates conflicting dilemmas. If decision-makers only sequentially select one of these decision paths, situations where norms and outcomes conflict would not arise in actual decision-making. Therefore, in situations with potential dilemmas, the decision processing is more likely to be parallel rather than sequential.

To address the aforementioned limitations, integrating the theoretical constructs of the CNI model, moral decision-making should comprehensively consider decision-makers' different response tendencies under the four combinations of the 2 (norm: prescribed/prohibited) × 2 (outcome: benefits greater than costs or costs greater than benefits) design. For example, in the CNI model' s situational structure, there are six scenario stories under each of these four combinations. Therefore, we can empirically calculate the probability of individuals choosing to accept action proposals in the six scenario stories under each combination. These four probability data points form a within-subjects 2 × 2 structure in the research design, allowing norms and outcomes to be treated as two parallel within-subjects factors in ANOVA designs. For between-group comparisons, grouping variables can serve as between-subjects factors. If norm or outcome factors show main effects, this indicates that decision-makers are driven by normative or consequentialist principles (i.e., the probability of accepting proposals when norms prescribe or benefits outweigh costs is significantly greater than when norms prohibit or costs outweigh benefits), or driven by anti-norm or anti-outcome principles (i.e., the probability of accepting proposals when norms prescribe or benefits outweigh costs is significantly smaller than when norms prohibit or costs outweigh benefits). If the interaction between norms and outcomes is significant, this indicates that the two principles do interact, and further simple effects analysis can identify the specific interaction pattern. Similarly, between-group effects and between-within interaction effects and their simple effects can all be examined. Notably, since the calculated data are probabilities of decision responses, Bonferroni correction is recommended for multiple comparisons.

Based on ANOVA and comprehensively considering decision-makers' potential parallel processing characteristics, we can also synthesize three dimensional indicators similar to the CNI model. [1] To distinguish it from the CNI model method and for mnemonic convenience when extracting the first letters of the dimensional indicators, we call this the CAN dimensional synthesis method:

Consequence sensitivity index C (Consequence sensitivity, C) = 1/2(p1 − p2 + p3 − p4); Norm sensitivity index N (Norm sensitivity, N) = 1/2(p3 − p1 + p4 − p2); Generalized action/inaction index A (generalized Action preference, A) = 1/4(p1 + p2 + p3 + p4).

The CAN dimensional synthesis method assumes from a parallel processing perspective that decision-makers simultaneously consider both norms and outcomes and are influenced by their interaction. If norms have a positive influence on decision-makers, then regardless of whether benefits outweigh costs or costs outweigh benefits, the probability of accepting action proposals when norms

prescribe should be significantly greater than when norms prohibit, with the difference representing norm sensitivity. When benefits outweigh costs, norm sensitivity can be represented by p3 − p1; when costs outweigh benefits, norm sensitivity can be represented by p4 − p2. Therefore, overall norm sensitivity can be represented by the average of these two: 1/2(p3 − p1 + p4 − p2). Similarly, if outcomes have a positive influence on decision-makers, then regardless of whether norms prescribe or prohibit, the probability of accepting action proposals when benefits outweigh costs should be significantly greater than when costs outweigh benefits, with the difference representing consequence sensitivity. When norms prohibit, consequence sensitivity can be represented by p1 − p2; when norms prescribe, consequence sensitivity can be represented by p3 − p4. Therefore, overall consequence sensitivity can be represented by the average of these two: 1/2(p1 − p2 + p3 − p4). As for the overall response tendency indicator, the average acceptance probability across all situations can reflect decision-makers' general preference: 1/4(p1 + p2 + p3 + p4).

For example, suppose an individual' s probabilities of choosing to accept action proposals across multiple scenario stories under the four combinations of norms and outcomes are p1=0.6, p2=0.1, p3=0.9, and p4=0.5. Then their norm sensitivity would be 1/2(p3 − p1 + p4 − p2)=0.35; their consequence sensitivity would be 1/2(p1 − p2 + p3 − p4)=0.45; and their overall preference would be 1/4(p1 + p2 + p3 + p4)=0.525. Similarly, these three indicator values can be calculated for each individual and then subjected to statistical testing.

In interpreting the indicators of the CAN dimensional synthesis method, if the C indicator is statistically significantly greater than 0, this indicates that decision-makers have a higher probability of accepting proposals when action outcomes produce benefits greater than costs compared to when they produce costs greater than benefits, thus being significantly driven by utilitarian (or consequentialist) principles. If significantly less than 0, this indicates significant anti-utilitarian (or anti-consequentialist) principle-driven behavior. If not significantly different from 0, this indicates that decisions are not driven by utilitarian principles. Similarly, if the N indicator is significantly greater than 0, this indicates that decision-makers are significantly driven by deontological (or normative) principles. If significantly less than 0, this indicates significant anti-deontological (or anti-normative) principle-driven behavior. If not significantly different from 0, this indicates that decision-makers are not driven by deontological principles. For the A indicator, if significantly greater than 0.5, this indicates a general preference for action over inaction. If significantly less than 0.5, this indicates a general preference for inaction over action. If not significantly different from 0.5, two situations may apply: (1) If both C and N indicators are also not significantly different from 0, this indicates that decision-makers are simply responding randomly; (2) If at least one of the C and N indicators is significantly different from 0, this indicates that decision-makers' overall general preferences for action and inaction are relatively balanced but are also influenced by utilitarian or deontological principles.

The CAN dimensional synthesis method is built upon the CNI model method, fully absorbing the theoretical construct advantages of the CNI model method, but the two differ in five aspects: First, the CNI model method assumes that decision-makers follow a sequential processing pattern of outcome→norm→general rejection/acceptance preference when making moral decisions, whereas the CAN dimensional synthesis method does not have such a priori assumptions but instead uses a common subtraction strategy in parameter synthesis, treating the effects of norms and outcomes equally and taking their average. Second, parameters derived from the CNI model method are at the group level and therefore cannot be used in correlational or regression designs; parameters can only be compared between two groups or against a specific value. In contrast, parameters obtained from the CAN dimensional synthesis method are at the individual level and can be used not only in correlational or regression designs but also for comparisons among multiple groups and for examining simple effects under ANOVA frameworks. Third, the I factor in the CNI model and the A factor in the CAN dimensional synthesis method have opposite statistical interpretations: larger values in the former represent stronger general rejection tendencies, while larger values in the latter represent stronger general acceptance tendencies. Fourth, the I factor in the CNI model represents decision-makers' general rejection/acceptance tendencies after stripping away the effects of outcomes and norms under the assumption of sequential processing, whereas the A factor in the CAN dimensional synthesis method represents decision-makers' overall acceptance/rejection tendencies without stripping away the effects of outcomes or norms, differing in connotation. Fifth, the CNI model method relies on a binary response mode where individuals must make binary judgments of either accepting or rejecting action proposals, whereas the CAN dimensional synthesis method does not have this requirement and can also be applied to continuous rating designs where decision-makers rate the degree of acceptance or rejection of action proposals to calculate corresponding parameters.

In summary, the CNI model method and CAN dimensional synthesis method can be used in combination and cross-referenced in actual research. In our research team, we have comprehensively used the aforementioned four research methods to explore moral question framing effects and have demonstrated the validity of the statistical test results of the CAN dimensional synthesis method, which has advantages such as allowing exploration of simple effects, being applicable to correlational and regression designs, and enabling multiple comparisons. [1]

## IV. Discussion and Outlook

Moral dilemma research is a typical representative of the broader social dilemma research, and its research methods have strong reference value and transferability for other social dilemma studies. Looking across the four empirical methods of moral dilemma research—from the classic dilemma method and process dissociation method to the CNI model method and the CAN dimensional synthesis

method proposed in this study—the theoretical constructs are downward compatible, with later methods encompassing the theoretical constructs of previous methods. Particularly, the CNI model method and CAN dimensional synthesis method have achieved comprehensive examination of all combinations of norms (prohibited or prescribed) and outcomes (benefits greater than costs or costs greater than benefits) and should be prioritized in future moral research, with potential for transfer to other social dilemma studies.

The CNI model method and CAN dimensional synthesis method have strong guiding significance for many previously unresolved controversies. First, the previously discovered positive correlation between psychopathy and utilitarian responses may exist only because psychopathy correlates with general rejection/acceptance tendencies, not necessarily with consequence sensitivity, which requires further testing in subsequent research. Second, previous research indicated that disgust makes moral judgments harsher, but recent studies show no such effect. [2] Additionally, whether emotional disgust arousal strengthens individuals' norm sensitivity or merely enhances their general rejection tendencies can also be answered using the latter two methods in this study. Third, whether there is a one-to-one correspondence between individuals' dual-process cognitive states and utilitarian/deontological tendencies also requires re-examination. Some scholars have proposed the possibility of intuitive utilitarianism, [3][4] which can also be tested using the latter two methods. Furthermore, recent scholars have used variants of the CNI model to examine the correspondence between norm/consequence sensitivity and deontology/utilitarianism, finding that parameters representing norms are also sensitive to outcomes. From this, they conclude that norms do not guide moral judgments unless they are expected to produce tangible results, suggesting that the division between norms and outcomes (or deontology and utilitarianism) as determinants of judgment is artificial. [1] This actually corresponds to the existence of interaction effects between norms and outcomes in the CAN dimensional synthesis method. Many other research controversies can also be more deeply addressed within the framework of this study.

The CNI model method and CAN dimensional synthesis method also have very strong theoretical and methodological expandability. As a theoretical and methodological approach, this method is not limited to moral dilemma research but can also be applied to other research topics with potential conflicts. Here are three examples. In determining moral culpability of actions, there has long been controversy between intention and consequence. Intentionalists argue that moral culpability lies in whether the intention involves subjective deliberateness, whereas consequentialists argue that moral culpability lies in whether the consequences involve harm or obstruction. Therefore, for determining moral culpability of actions, intention and consequence form a potential conflict, with possible situational combinations including deliberate and harmful, deliberate and harmless, non-deliberate and harmful, and non-deliberate and harmless. These can be studied using the theoretical constructs of the CNI model method and CAN dimensional synthesis method. Recent scholars studying emotional responses to

harm have also recommended this further testing approach. [2] Similarly, in personal moral preferences and moral choices, traditional Chinese culture has long debated between public justice and private interest, which often create potential conflicts, with situational combinations including benefiting public justice and private interest, benefiting public justice while harming private interest, harming public justice while benefiting private interest, and harming both public justice and private interest. These can also be analyzed using similar methods. Furthermore, in consumer decision-making, the cost-performance issue between performance and price is also a common contradiction, with possible combinations including high performance and high price, high performance and low price, low performance and high price, and low performance and low price. These can also be deeply explored using the aforementioned methods. Therefore, beyond advancing moral dilemma research, the CNI model method and CAN dimensional synthesis method have cross-disciplinary and cross-domain applicability. Any research topic with potential conflicts can consider using a similar methodological framework for investigation.

In summary, moral dilemma research has undergone methodological evolution through four approaches: the classic dilemma method, process dissociation method, CNI model method, and CAN dimensional synthesis method. Theoretical constructs have developed from single-type conflict situations to comprehensive examination of all four combinations of norms (prohibited or prescribed) and outcomes (benefits greater than costs or costs greater than benefits). Particularly, the CNI model method and CAN dimensional synthesis method provide methodological approaches for resolving many controversies in previous research. Moreover, this framework is not limited to moral dilemma research but also provides methodological references for cross-disciplinary and cross-domain social dilemmas and other topics with potential conflicts.

## ABSTRACT

Social dilemma is a multi-disciplinary topic widely discussed in ethics, psychology, sociology, and economics, with moral dilemma as its typical paradigm and a long research history. The present article reviews three existing empirical approaches in dilemma research: the traditional dilemma paradigm, process dissociation method, and Consequences-Norms-generalized Inaction/action (CNI) model. Based on an analysis of the contributions and limitations of these three approaches, we developed a Consequences-general Action/inaction preference-Norms sensitivity estimation (CAN) algorithm. With the development of these four approaches, moral dilemma research is no longer limited to contradictory dilemma situations. The CNI model and CAN algorithm have extended to consider the four combinations of proscriptive/prescriptive norms and benefits greater/smaller than costs. With the CNI model and CAN algorithm, controversies in moral theories and empirical inconsistencies can be further clarified. These four approaches provide methodological references for similar topics with potential contradictions in many other domains and can be used across multiple

disciplines.

**Keywords:** empirical approaches; social dilemma; moral dilemma; conflict rules

**About the Authors:** Liu Chuanjun, Ph.D. candidate in Social Psychology, Department of Psychology, Tsinghua University, intermediate professional title, Tel: 13550836591, Email: lcj17@mails.tsinghua.edu.cn, Address: Weiqing Building, Department of Psychology, School of Social Sciences, Tsinghua University, Qinghuayuan 1, Haidian District, Beijing, 100084.

Liao Jiangqun, Ph.D., Associate Professor, Department of Psychology, Tsinghua University, doctoral supervisor, Tel: 010-62787208, Email: liaojq@tsinghua.edu.cn, Address: Room 210, Mingzhai, School of Social Sciences, Tsinghua University, Qinghuayuan 1, Haidian District, Beijing, 100084.

*Note: Figure translations are in progress. See original paper for figures.*