WHEN YOUR LOVED ONE IS ON THE TROLLEY TRACK: A STUDY OF MORAL DILEMMA WHEN CLOSE RELATIONSHIP IS INVOLVED

A Thesis

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Master of Arts

by

Bian Deng

August 2019

ABSTRACT

The present study investigated how Chinese and western people would response to hypothetical and real-life moral dilemmas with different consequences and numbers of victims when the one they care about is involved. We used online surveys to collect responses from both Chinese and western participants, and explored the threshold of how many strangers (or their benefits, depending on the scenarios) people are willing to sacrifice for the benefit of their loved ones. We concluded that people are willing to sacrifice more strangers (or their properties) to save the life (or the properties) of the one they care about, comparing to the amount of strangers they are willing to leave emotionally distressed to make the one they care about avoid being distressed. Chinese participants have higher thresholds than western participants in such situations, and we identified 4 major moral principles behind people's moral dilemma decisions.

BIOGRAPHICAL SKETCH

Bian Deng graduated with a Bachelor's degree in the School of Psychology and Cognitive Science from Peking University, China. During his undergraduate study, his research explored a wide range of children's development, including infants' understanding of mutuality in social interaction, and the role of Chinese classifiers in preschooler's quantification. He is currently working on his Masters in Human Development at Cornell University, extending his research interests into people's moral decision-making, especially in real life situations.

ACKNOWLEDGMENTS

I would like to express my sincere appreciation to my major advisor Professor Stephen Ceci for his tremendous supports and insightful suggestions throughout this research. I would not be able to complete the research without his consistent encouragement and guidance.

I am also grateful for the help from Professor Felix Thoemmes, especially for giving me advice and instructions in data analyses, and the extended support after the thesis defense.

I would also like to thank my fellow lab member Vivian Rotenstein, along with my Research Assistants, for helping me construct the survey and collecting data from western participants.

TABLE OF CONTENTS

1 Introduction

The daughter of electronic repair shop owner Daisuke Endo was critically ill and in need of a heart transplant. However, she happened to possess a rare blood type which made it very difficult to find a heart donor. One day his lifelong friend Tateno, a local policeman, went out on a mission to protect a girl called Hitomi, who was scheduled to hand her sister's ransom to a kidnapping gang. It looked like a normal mission for a well-established detective, but there was only one problem: Hitomi was known to share the same blood type with Endo's daughter. After the ransom handoff, Tateno's colleagues pursued the suspect, leaving him alone with the innocent girl, Hitomi. Realizing it was probably the only chance to save his friend's daughter whom he considered as dear as his own, Tateno slowly raised his gun…

This is the beginning of a famous mystery novel "428: Shibuya Scramble". Tateno was facing a painful moral dilemma: to protect this innocent girl, or to kill her and save his friend's daughter whom he considered his own daughter.

One of the most best-known moral dilemmas is the so-called "trolley car problem". The original depiction was that there were 5 people tied on a trolley car track and a car was about to run over them. Near the track there was a lever which can be used to direct the car to a side track so the 5 people would be saved. However, another person was tied on the side track so if the lever was pulled he would be run over instead. A multitude of variations have been proposed since, and a popular explanation put forward in resolving such moral dilemmas is utilitarianism, which emphasizes the maximum benefits of a

behavior without considering deontological considerations and other restrictions, as long as the most benefit can be obtained for the most people in the end (Cao, et. al, 2017).

However, in real life situations, such thinking may not always work out, as the assumption “without considering other restrictions” is not readily satisfied in various contexts. What’s more, there is experimental evidence that utilitarian thinking does not necessarily lead to such behaviors or actions (Tassy, et. al, 2013; Gold, Coleman, & Pulford, 2014). Particularly, in Tateno’s case, it was the relationship between him and one of the potential victims that tipped the balance, even apparently abandoning his duty as a detective.

Tateno’s story leads us to the intriguing question: how would people behave in real-life moral dilemmas when people they care about are involved? In past research about moral dilemmas, the potential “victims” are usually anonymous and unknown to the individual making the decision. From the perspective of experimental design, the use of anonymous characters serves to control potential confounds, and it also creates a perfect setting where no other restrictions are included. However, when the personal relationship between the decision-maker and the potential victim is taken into consideration, as we’ve seen in Tateno’s case, people’s behavior can be drastically affected, even to the point where it violates their other moral principles. We’ve seen parents that are willing to “sacrifice anything” to protect benefits to their children, even if that means they would have to go out of their way to hurt, even kill somebody else. (A less extreme example might be the current scandal surrounding college

admission bribery at select universities.) Would people's desire to save their loved ones outweigh a traditional utilitarian calculus? What if the potential sacrifice is even bigger than 5 other people? In our current study, we varied the number of strangers that would potentially suffer in the trolley car dilemma, from 1 to 1000, to see how far people are willing to go for the benefit of their loved one.

In real life situations, moral dilemmas may appear in a wide range of contexts, with various things at stake. Although they could involve a life-threatening situation, more often they would involve some less extreme kinds of losses such as financial loss or emotional distress. Gold, Pulford and Coleman (2013) designed an experiment to compare people's choices on moral dilemma questions involving death, job loss, financial loss and emotional stress. They found that people believed it more acceptable to apply the utilitarian principle in job loss and financial loss situations than in emotional distress situations, but, rather surprisingly, not significantly different from death situation. However, we haven't found any previous studies that have explicitly explored how people would respond to real-life moral dilemmas when close relationships were involved. The present research was designed to fill this void.

An additional goal of the present research was to examine the potential cultural differences between western and eastern participants in moral dilemma responses. Several previous works had noticed cultural differences when it comes to moral decisions (Miller & Bresoff, 1998; O'Neill & Petrinovich, 1998). More specifically, a cross-cultural study by Gold, Colman and Pulford

(2014) compared British and Chinese participants' behaviors and judgments on moral dilemmas, which also included non-death real-life situations. Their findings suggest that Chinese participants are less willing to sacrifice one to save five. However, their findings also suggested that the result could possibly be due to Chinese people possessing a stronger tendency to "not intervene" rather than any deontological principle. In the current study, we constructed real-life scenarios in which a clear decision has to be made to decide who (or whose benefit) will be preferenced, in an effort to eliminate people's notion of "not doing anything".

2 Methods and Design

Participants. Participants were voluntary respondents to an online survey. Western participants were recruited from online MTurk survey, Eastern participants were recruited at Peking University in China.

Design and materials. We used a mixed design. Each subject was asked to judge 12 scenarios in a 3(Scenarios: trolley vs. emotional distress vs. financial loss) * 4(Number of strangers, 1 vs. 10 vs. 100 vs. 1000) within-subjects design, and the two cultural differences (western vs. eastern) is a between-subject factor.

The participants were asked to complete a survey containing 3 sets of

moral dilemma questions on a Qualtrics survey platform (a Chinese translated version of this survey was distributed to Chinese participants), each set consists of 4 questions that varies in how many strangers were involved in the dilemma.

The Trolley car scenario:

Think about the person you love and care most deeply about.

Now, imagine that you see a runaway trolley car speeding towards this person, who is incapacitated on the tracks.

You are standing next to a lever which controls the direction in which the trolley car is headed.

On a side track, you see 1/10/100/1000 *stranger(s), lying down tied up on the tracks.*

If you pull the lever, the train will be redirected to the stranger, saving the person you care about, and striking the stranger.

If you do not pull the lever, however, the stranger will be saved, and the person you love will be struck by the trolley.

The emotional distress and financial loss scenario both featured a situation in which the participant had to choose his/her children's (which was supposed to be the equivalent of "someone you care most deeply about, in order to make the real-life scenario easier to imagine) interest or some strangers'.

The emotional stress scenario:

Suppose you're driving a van every morning that would take 1/10/100/1000 *adult(s) to their workplace, and your child to school.*

Normally you could make it to both places on time, regardless of the sequence of destinations you go to.

However, on this particular morning, the traffic is extremely heavy, and you estimate that either your child or the 1 adult will be late (depending on which destination you go first), and because all of you value being on time, being late would cause extreme emotional stress.

The financial loss scenario:

Suppose you're a janitor who is also serving as a security guard in a gym locker room.

During your shift, you see your child putting his smartphone and laptop in a locker on the right side of the gym, and 1/10/100/1000 *stranger(s) put theirs smartphone and laptop in a locker on the left side of the gym.*

As you're cleaning, both lockers start to fall down.

You can only stop one locker from falling over, but the other locker will collapse, thus breaking all the valuable items stored inside.

Participants response to the question of whether pull the lever or not in the trolley car scenario, or whether prioritizing their child's interest or the strangers' were the dependent measures. Their responses and the time they took were recorded through the insertion of timers on each decision.

The response were recorded as a binary variable indicating "saving the one you care for" or "saving the number of strangers". We hypothesized that generally, as the number of strangers increases, people are more likely to choose "saving strangers" over "saving the one you care for", and a "threshold" exists for every people in each scenario. For example, if a particular participant chose to save the one he cared for against 1- and 10-strangers but otherwise in 100- and 1000-stranger scenarios in Trolley car situation, his "threshold" in

Trolley car situation would be somewhere between 10 and 100 strangers. In this sense, the number of strangers also serves as a scale for “how much people value the interests (lives/avoiding emotional distress/avoiding financial loss) of the one they care about”.

These response latencies could be used to infer implicit processing that might reflect competing values that might be obscured by relying exclusively on the final decision as is commonly done in trolley car studies. We hoped that this dependent measure would add sensitivity because it is possible that with enough time to carefully consider a scenario one’s true feelings may emerge. In contrast, with ample time to consider the scenario participants may abandon their true desires and respond in a socially acceptable manner. Thus, the timers give a measure of how troubling they find the dilemma.

3 Results

3.1 Descriptive statistics

We’ve collected 131 responses from the Chinese data, 17 of which were invalid due to either misunderstanding the question or missing data, resulting in 114 valid data (Male=54, Female=60).

The western data, collected via MTurk, consisted a total of 58, 9 of which were invalid due to failing the attention-checking question, and 2 of the remaining had duplicated MTurk ID (A2AJT6HM3UFTFB) but with different responses, so we were unable to determine which set of data should be valid, thus we considered both sets of data invalid, resulting in 47 valid data (Male=30, Female=17).

We recorded the response of "aiding the one you care about" as 0, and the other option (aiding the strangers) as 1. The overall responses are shown in Table 1 below.

Table1: Overall counts of responses among all participants

Scenario	Number of strangers	Chinese		Western	
		Save Love	Save Stranger	Save Love	Save Stranger
Trolley Car	1	105	9	43	4
	10	66	48	33	14
	100	41	73	26	21
	1000	40	74	26	21
Emotional Distress	1	54	60	27	20
	10	21	93	13	34
	100	8	106	12	35
	1000	11	103	13	34
Financial Loss	1	98	16	35	12
	10	66	48	21	26
	100	45	69	21	26
	1000	37	77	20	27

The secondary variable we assessed, the time taken to make the decision, was also recorded. However, we noticed there are several long response latencies, especially in the first question of each series. From the feedback we've collected from participants we noticed some of them said that during the first question they were still processing the scenario, thus took longer time (although in our survey design we did separate the scenario and question in an effort to reduce such confound). We intended to limit the participants' response time within 5 seconds, but we would rather have a delayed response than have a missing value after 5 seconds expired. In this case, we decided to record the responses that took more than 5 seconds as 5 seconds, and the statistics were shown in the following table.

Table 2: Response time (sec)

Scenario	Number of strangers	Chinese		Western	
		Mean*	SD	Mean	SD
Trolley Car	1	5.22(4.32)	2.63(0.97)	4.63(3.81)	2.61(1.10)
	10	3.96(3.38)	2.87(1.36)	4.66(3.74)	3.53(1.27)
	100	3.57(2.83)	4.37(1.47)	4.35(3.46)	4.14(1..41)
	1000	3.08(2.76)	2.71(1.30)	3.77(3.31)	2.43(1.35)
Emotional Distress	1	4.10(3.34)	3.29(1.28)	5.51(3.83)	6.52(1.05)
	10	2.25(2.24)	1.29(1.28)	3.12(3.01)	1.77(1.49)
	100	1.85(1.85)	1.13(1.12)	2.51(2.48)	1.33(1.28)
	1000	1.84(1.83)	1.14(1.13)	2.82(2.58)	1.94(1.27)
Financial Loss	1	3.32(3.02)	2.34(1.13)	3.78(3.11)	4.65(1.36)
	10	2.79(2.32)	2.91(1.37)	3.51(2.65)	4.04(1.34)
	100	1.92(1.92)	1.18(1.17)	2.45(2.36)	1.48(1.16)
	1000	1.90(1.90)	1.18(1.17)	2.41(2.20)	2.10(1.16)

* original stats (adjusted stats)

3.2 Response

The response data, due to its binary nature, is then processed in a Wald χ^2 test to examine the main effects and interactions.

Table 3: Wald χ^2 test of the model

Factor	Wald χ^2	*df*	*P*
(Intercept)	1.263	1	.261
Scenario	75.465	2	<.001
Number	140.771	3	<.001
Culture	5.187	1	.023
Scenario * Number	10.383	6	.109
Scenario * Culture	8.955	2	.011
Number * Culture	10.279	3	.016
Scenario * Number * Culture	3.493	6	.745

The result shows all 3 main effects are significant, and interactions between culture and both within-subject factors are significant. The interaction between Scenario and Number was not significance, nor was three-way interactions. We proceeded to check pairwise comparisons among interactions (Bonferroni correction was used in all pairwise comparisons).

3.2.1. Overall Scenario

First we examine the pairwise differences of numbers of strangers within each scenario. The overall effect averaged across culture is shown in Table 4 below.

Table 4: Pairwise comparisons between different numbers of stranger in each scenario, averaged across culture.

Scenario		Number difference and *p*-value				
		Mean	1	10	100	1000
Trolley	1	.08		<.001	<.001	<.001
	10	.36	-.28		<.001	.003
	100	.55	-.46	-.19		1.000
	1000	.55	-.47	-.19	.00	
Emotional	1	.48		<.001	<.001	<.001
	10	.77	-.30		.517	1.000
	100	.86	-.39	-.09		1.000
	1000	.83	-.36	-.06	.03	
Financial	1	.19		<.001	<.001	<.001
	10	.49	-.30		1.000	.024
	100	.58	-.39	-.09		1.000
	1000	.63	-.44	-.14	-.05	

The difference in the table is always "the less stranger group-the more stranger group", bottom-left: mean difference; top-right: *p*-value.

The mean values of each group (based on our coding this is essentially the ratio of people who chose to help strangers) give us inferences on where do the "thresholds" of the participants lie. For example, the mean value of Trolley car-1000 stranger group is .55, meaning that 45% of the participants have their threshold in such situation as "more than 1000" (willing to sacrifice more than 1000 to save the one they care for). The difference between each two number groups should suggest the proportion of participants who have the threshold between those two numbers.

People are much more willing to sacrifice 1 stranger (or their benefits) to preserve that of someone they care about, comparing to 10 or more strangers. The difference between 1- and 10-stranger situations show that there are approximately 30 percent of the participants that have their thresholds between 1 and 10 strangers in each of the three scenarios. Also, there's no significant difference between 100 and 1000 strangers in each of the three scenario groups, which means almost nobody has their threshold at 100 and 1000. In addition, only in Trolley car scenario there is significant difference between 10- and 100-stranger situations. Such results will provide important implication on the position of thresholds in each different context.

3.2.2. Overall Number of strangers

Similarly we examine the pairwise differences of scenarios within each number group. The overall effect averaged across culture is shown in Table 5 below.

Table 5: Pairwise comparisons between different scenarios in each number, averaged across culture.

Number		Scenario difference			
		Mean	Trolley	Emotional	Financial
1	Trolley	.08		<.001	.445
	Emotional	.48	-.39		<.001
	Financial	.19	-.11	.28	
10	Trolley	.36		<.001	1.000
	Emotional	.77	-.42		<.001
	Financial	.49	-.13	.29	
100	Trolley	.55		<.001	1.000
	Emotional	.86	-.32		<.001
	Financial	.58	-.03	.28	
1000	Trolley	.55		<.001	1.000

Emotional	.83	-.28		.002
Financial	.63	-.08	.21	

The difference in the table is always "the less stranger group-the more stranger group", bottom-left: mean difference; top-right: *p*-value.

The 1-stranger scenario can be considered as a "baseline point", as there is no difference in number of people involved on both sides, which is the traditional utilitarian criterion. At this point, we noticed significant higher baseline between Emotional distress and other two scenarios, and such pattern remains at each level of number. A little bit surprisingly, no significant difference was found at any single level of number between Trolley car and Financial loss scenarios.

3.2.3. Cultural difference

Because of the significant interaction, we examine the difference in conditional effects when culture group is held constant.

Table 6: Pairwise comparisons between different numbers of stranger in each scenario, split by culture

Scenario	Number difference (*p*-value)				
		1	10	100	1000
Trolley	1		-.21(1.000)	-.36(.003)	-.36(.003)
	10	-.34(<.001)		-.15(1.000)	-.15(1.000)
	100	-.56(<.001)	-.22(<.001)		.00(1.000)
	1000	-.57(<.001)	-.23(<.001)	.00(1.000)	
Emotional	1		-.30(.046)	-.32(.005)	-.30(.046)
	10	-.29(<.001)		-.02(1.000)	.00(1.000)
	100	-.40(<.001)	-.11(.112)		.02(1.000)
	1000	-.38(<.001)	-.09(1.000)	.03(1.000)	
Financial	1		-.30(.120)	-.30(.120)	-.32(.054)
	10	-.28(<.001)		.00(1.000)	-.02(1.000)
	100	-.46(<.001)	-.18(<.001)		-.02(1.000)
	1000	-.54(<.001)	-.25(<.001)	-.07(1.000)	

The difference in the table is always "the less stranger group-the

more stranger group", bottom-left: Chinese; top-right: Western.

Among Chinese participants, significant differences between 1-stranger and each of the other three number levels are found in all three scenarios. In addition, the differences between 10- and 100-strangers in both Trolley car and Financial loss scenarios are also significant.

However, among western participants the trends are different. No significant difference is detected among 10-, 100- and 1000-stranger groups, regardless of scenario. In the Trolley car scenario, significant difference is observed between 1- and 100(as well as 1000)- stranger groups; whereas in Emotional distress scenario 1-stranger group is significantly different from all 3 other number levels. No significant difference is detected among western participants across 4 number levels in Financial loss scenario.

3.3 Time

The time data is examined in a repeated measures test in SPSS. We hypothesized that, in each scenario, it would be tougher to make decision when the number of strangers are close to the threshold. If the number of strangers are far less than the threshold it would be relatively easy to sacrifice those strangers, and when the number of strangers increased past the threshold, it should be easier to determine that they should save the large number of strangers, which means they may take less time to make decisions.

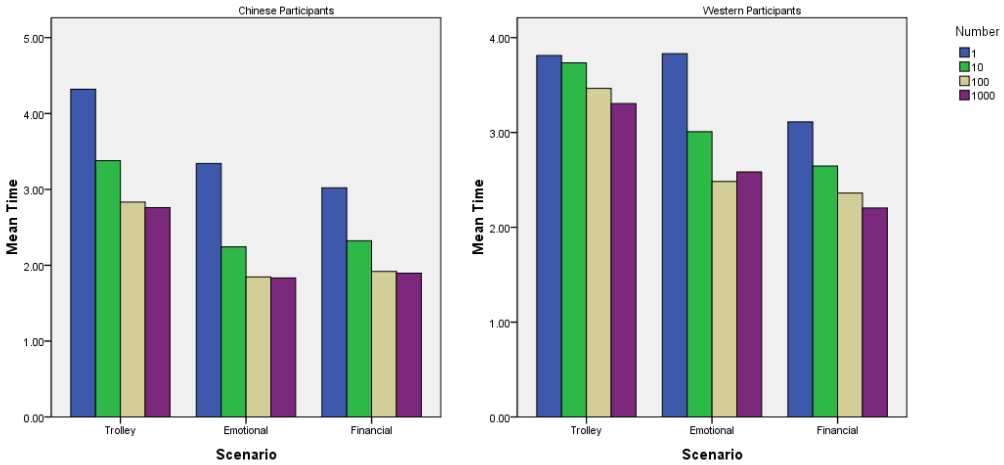

Figure 1: Mean adjusted time taken to make decision (Left: Chinese, Right: Western)

Table7: Repeated Measures Test result of Time

Factor	*F*	*df*	*p*-value	Partial η^2
Scenario	95.640	1.943	<.001	.376
Number	111.814	2.904	<.001	.413
Scenario * Number	2.278	5.871	.036	.014
Culture	9.586	1	.002	.057
Scenario * Culture	4.190	1.943	.017	.026
Number * Culture	6.553	2.904	<.001	.040
Scenario * Number * Culture	1.996	5.871	.065	.012

Note: Huynh-Feldt correction are performed in all within-subject variables due to violation of sphericity.

However, as the graph shows, even after the adjustment for long response latencies, the longest response time in each scenario almost always occurred in the "1 stranger" case, and the time taken generally decreased with larger numbers. Despite this unexpected effect, we were still able to identify significant differences between Chinese and western participants (F=9.586, p<.01), which showed that Chinese participants took less time on average than western participants. We also found a significant interaction between Scenario and Number (F=2.278, p<.05, Huynh-Feldt correction). Most notably, we found no difference in response time between 100- and 1000-stranger situation

in all scenarios, and a significant time decrease from 10 to 100 strangers, supporting our inference of thresholds from response data.

In addition to the overall average pattern we are also interested in how each participant's time taken changes when the number of strangers increases in each scenario. The majority of participants took the longest time in the 1-stranger condition (72.7% in Trolley car, 73.3% in Emotional distress and 59% in Financial loss), again in accord with the overall result. However, because of the relatively large distance between the options, it's difficult to observe distinct patterns in individual levels, or to examine our hypothesis that the longer time would be taken if the number of strangers is closer to the threshold.

We also conducted more detailed pairwise comparisons like we did in the Response part.

3.3.1. Overall scenario

In each scenario, The time taken by participants demonstrated a decreasing pattern (as shown in Table 8), in which 1-stranger condition took the most time, 10-stranger condition took significantly less time, and again less time taken when it came to 100 or more strangers. The result suggests that people are having a much harder time deciding the fate of less people and, according to our hypothesis, indicates more people would have their threshold at lower number of strangers (closer to 1-10 strangers instead of 100 or more).

Table 8: Pairwise comparisons of time between different numbers of stranger in each scenario, averaged across culture.

Scenario	Time difference and *p*-value

		Mean	1	10	100	1000
Trolley	1	4.07		<.001	<.001	<.001
	10	3.58	.50		.006	<.001
	100	3.15	.92	.41		1.000
	1000	3.03	1.03	.53	.11	
Emotional	1	3.59		<.001	<.001	<.001
	10	2.63	.96		.001	.002
	100	2.16	1.42	.46		1.000
	1000	2.21	1.38	.42	-.04	
Financial	1	3.07		<.001	<.001	<.001
	10	2.49	.58		.042	.005
	100	2.14	.93	.34		1.000
	1000	2.05	1.02	.44	.09	

The difference in the table is always "the less stranger group-the more stranger group", bottom-left: mean difference; top-right: *p*-value.

3.3.2. Overall number of strangers

We also compared the time taken in different scenarios within each level of number of strangers. We found that people spend more time in Trolley car scenario than the two real-life scenarios in each number group. The result indicates that people might generally find it tougher if a moral dilemma involves lives and deaths than those real-life situations where emotional or financial interests are at stake. In most cases there were no significant differences between Emotional distress and Financial loss situations (only difference was observed in 1-stranger condition).

Table 9: Pairwise comparisons of time between different numbers of stranger in each scenario, averaged across culture.

Number		Time difference			
		Mean	Trolley	Emotional	Financial
1	Trolley	4.07		<.001	<.001
	Emotional	3.59	.48		<.001
	Financial	3.07	1.00	.52	

10	Trolley	3.58		<.001	<.001
	Emotional	2.63	.93		.858
	Financial	2.49	1.07	.14	
100	Trolley	3.15		<.001	<.001
	Emotional	2.16	.98		1.000
	Financial	2.14	1.01	0.2	
1000	Trolley	3.03		<.001	<.001
	Emotional	2.21	.82		.523
	Financial	2.05	.98	.16	

The difference in the table is always "the less stranger group-the more stranger group", bottom-left: mean difference; top-right: *p*-value.

3.3.3. Cultural comparisons

We compared cultural difference of time taken both by scenario (Table 10) and by number of strangers (Table 11).

In scenario comparisons Chinese participants made decisions significantly faster in Emotional distress scenario ($p<.001$) and the difference in Financial loss scenario reached borderline significance ($p=.051$). The time difference in Trolley car scenario is not significant.

Table 10: Cross-cultural comparison of Time by Scenario

Scenario	Culture	Mean	Mean dif.	*p*-value
Trolley	CN	3.323	.256	.130
	West	3.579		
Emotional	CN	2.315	.662	<.001
	West	2.977		
Financial	CN	2.289	.292	.051
	West	2.581		

In number comparisons Chinese participants made decisions significantly faster in all but the 1-stranger condition, in which the difference is not significant.

Table 11: Cross-cultural comparison of Time by Number

Number	Culture	Mean	Mean dif.	*p*-value
1	CN	3.561	.024	.870
	West	3.585		
10	CN	2.648	.482	.005
	West	3.130		
100	CN	2.199	.571	<001
	West	2.770		
1000	CN	2.163	.535	<.001
	West	2.698		

However, we are aware that the difference between conditions could be confounded by the sequence of presentation. We didn't randomize the sequence of questions; instead every set of question is presented in the order of ascending number of strangers. Participants might be setting up their "criteria" while answering the first questions, and apply those criteria when answering later questions to more efficiently make decisions. It was also possible that they were able to anticipate the pattern of questions so they needed less time to make decisions when the questionnaire progressed. The cultural differences were insignificant in the "first" conditions of both scenario and number, which also possibly indicates that those Chinese participants were faster to establish a decision-making pattern.

3.4 Reasoning

So, what are the rationales behind their decisions, or in other word, how do people determine the thresholds? To answer this question, we have to look into the write-ups participants had provided, and we roughly divide the responses into 3 groups by their threshold: "Altruistic group", in which they chose to help even 1 stranger over the one they care about; "Selfish group", in

which they always chose the one they care about no matter the number of strangers, and “Balance group”, in which they initially chose to help their loved ones against 1 stranger but changed their minds somewhere in between. The prevailing reasons in each group are different. We generalized 4 major categories of moral reasoning based on their write-ups:

(1) Recovery principle. When this principle is applied, people concede a loss that they believe can be recovered, or is easier to recover compared with the other alternative;

(2) Responsibility principle. When this principle is applied, people choose the action they feel obliged to do, or choose against what they don’t feel obliged to do;

(3) Selfish principle. When this principle is applied, people solely focus on the benefit they would get themselves without even considering what loss other people might suffer;

(4) Pseudo-utilitarian principle. When this principle is applied, people acknowledge the number of strangers that would be sacrificed is important, and believe that when the number of innocent sacrifices is large enough then they should offset the personal interest of saving the one they care about.

Besides the 4 major principles, there are also some other replies that didn’t fit into any of them, mostly being too specific and scenario related. For example, “I want to make my child know the importance of punctuality (so I would drive the adults to work first)”, “It’s hard to punish a lot of people for being late at the same time (so I would drive my child to work first)”. Also

there were some vague answers which didn't really explain the reasoning, such as "I trust on my instinct", which were left out in our analysis.

In the "Altruistic group", which is almost exclusively seen in the two real-life scenarios, we identified Recovery and Responsibility as 2 major rationales that lead to their decisions. Answers like "I can console the child who suffers emotional distress", "I can always buy a new smartphone for my child", or "Being late for school is not as serious as late for work" were coded as Recovery principle being applied, since they focused on the ability to redeem the loss. Answers like "Sending the adults to work is my duty" were categorized as Responsibility principle being applied.

For the "Selfish" group, most people just simply stated the "selfish" principle, especially in the Trolley car scenario. Among the participants' write-ups, one Chinese participant quoted the famous sentence in *The Tale of Three Kingdoms*, "I would rather betray everyone than being betrayed by anyone," to describe the reason to sacrifice 1000 strangers to save the one he care about. In our scenarios, any reasoning that implicated the significance of close relationship is considered "selfish" principle being applied. Other reasons in this category include "No reason is needed if it's for the life of my loved one", "I don't care about strangers' benefit", or "I put my emotions and feelings at the highest priority". Besides, Responsibility rationale is also observed among this group. If the participants interpret the situation as they are not responsible for the loss of whatever number of strangers' properties, they don't feel obliged to try to protect it hence justified their decision to protect the properties of the

one person they care about. Reasons like “It’s not my fault that the strangers’ properties were damaged”, “I’m not responsible for the loss of strangers” would fit into this category.

Finally, for all those in “Balance” group, there were no such clear-cut answers. The most general answer was that they “wanted to minimize the total loss/maximize the total benefits”, or “the number of strangers matters”, somewhat similar to the utilitarian thinking. Reasons that highlighted this line of thinking included “I only had to suffer a small loss (financially) to save the majority”, “Comparison of the value (of items in jeopardy)”, or “I would reconsider if I would have to sacrifice a ‘considerable amount of strangers’ in order to save my loved one”. We named it “Pseudo-utilitarian” because it resembles the utilitarian idea but doesn’t necessarily assume that all individuals are equal. The rule of Recovery can also play a significant role in this group under non-lethal scenarios, as some people justified their selfish decision when less strangers’ properties were jeopardized, because such loss is easier to recover than public loss on a larger scale. This rationale is particularly identified in some cases when people chose to sacrifice the interests of 10 strangers but favored 100 and more strangers, when they believed that it would no longer be affordable if the loss is that much.

We counted how many times people cited the aforementioned principles in their responses. In some cases people used multiple principles to explain their decision in one single condition, and in such cases they were counted in both categories.

Table 12: Occurrence of each principles in each condition

Scenario		Reasoning				
		Rec	Res	Pseudo-Uti	Self	Other
Trolley	Selfish	0	0	0	45	6
	Balanced	0	0	71	0	17
	Altruistic	0	0	0	0	14
Emotional	Selfish	0	0	0	7	2
	Balanced	19	0	42	0	19
	Altruistic	35	29	0	0	15
Financial	Selfish	0	32	0	16	3
	Balanced	6	3	55	0	11
	Altruistic	9	2	3	0	7

(Chinese and Western combined.)

4. Discussion

We intended to explore how people would response facing hypothetical and real-life moral dilemmas when someone they care deeply about (their loved ones) is involved. From the responses they made, we are able to infer how much do people value their loved ones on the scale of “number of strangers” when different contexts are presented, and, along with the reasoning they provided in the write-ups, also the moral principles that help them make the decisions.

Overall, people are willing to go further to save the lives of those they care about than only to help them avoid emotional distress or financial loss. 45% of our participants felt it worth to sacrifice even 1000 innocent lives to save the life of one person they care about, while only 23% believed that they should help their child avoid extreme emotional distress at the expense of “only” 10 strangers suffer the same pressure. It’s quite shocking to see that such a large proportion of the participants are willing to “kill” thousands of people to save

one person they care about.

When we look at the threshold of how many strangers could outweigh the one they care about, apart from those who has the threshold over 1000 and under the baseline, the most significant shift in each group are all in 1-10 strangers. Roughly 30% of participants in each scenario who had prioritized the one they care about over 1 stranger changed their mind when the number of strangers rose to 10. There were essentially no difference between 100 and 1000 groups, which indicates that if the sacrifice goes to more than 100, the number hardly matters anymore.

The cultural difference mainly displays itself in the 10-100 strangers interval. Western participants showed no significant difference between these two groups in all scenarios, while responses from Chinese participants significantly differed in both Trolley car and Financial loss scenarios. It means there are more Chinese who have their threshold between 10 and100, possibly indicating that Chinese are more likely to assign higher value to the interests of the one they care about than the western counterparts.

We also assessed the time taken to answer each question. The no-difference between 100- and 1000-stranger groups partially supported our inference of thresholds from the response data. However, our assumption of the time should be longer when the number of strangers are closer to the threshold may not hold true. First, there's significantly longer time taken in 1-stranger situation, which is the first question of each scenario. This can be confounded by people still digesting the plot of the story (despite we made effort to prevent

that from happening), or maybe they are actually establishing a judgment criteria for the ongoing scenario during that first question. The consistently decreasing pattern in response time could imply that participants in later questions had already set up their decision criteria and threshold before they enter the choice page, so they might not feel that difficult even if they are crossing their threshold.

To summarize people's reasoning behind moral decisions, we identified 4 major moral principles that govern people's decision in moral dilemmas when someone they care about is involved: "Recovery" (whether the punishment is redeemable), "Responsibility" (whether they are obliged to do one thing over the other, and/or would they be held responsible for whatever loss occurs), "Selfish" (bias towards the one they care about no matter what the cost and consequences are), and "Pseudo-utilitarian" (maximize the total loss, but from a biased personal perspective which assigns extra weigh on the interests of the one they care about). Under different circumstances and across different threshold groups, people use different principles to make (and justify) their decisions, especially in real-life situations where more people are using Recovery and Responsibility principles.

5. Limitations and Future directions

Because of the lacking of previous research on moral dilemmas with both real-life situations and close relationship involved, our current research have few examples to follow, which created some obstacles in our experiment design.

The most significant difficulty we encountered is to invent the real-life scenarios. The study of Gold et. al. (2013) invented a “sauna scenario”, in which the people in a sauna room could face punishment for them being exposed on a public camera. In their material they were able to control all other variables except for the would-be punishment. However, their scenarios were simulating the bystander and footbridge models of trolley car, which means they cannot prevent the option of “not taking action” which we intended to eliminate. In addition, we want the punishment in our scenarios more “natural”, and we feel that the “sauna scenario”, it is unnatural for the people in the story to get punished when the party at fault is actually the sauna house which didn’t fix a malfunctioning hidden camera in the sauna house and didn’t inform the customers. These are the rationale behind our “drive to work” and “falling lockers” stories. However, in our real-life scenarios, we used “your child” as “the one you cared deeply about” and “adults” as the strangers, that creates an unnecessary discrepancy. Despite we explicitly described the would-be punishment in each of the occasions, some people still were inclined to think one adult should outweigh one child because of the difference in their social status. This is probably the most significant confound in our material and should be fixed in our further studies. A potential remedy is to make refinement on the “drive to work” story, change the “drive your child to school” into “drive your spouse (or other adult that you care deeply about) to work”. Such refinement can not only help eliminating the child-adult discrepancy, but also adapt well in financial loss (receive cut in salary if late) and some other real-

life scenarios.

In our research, the 1 – 1000 scale of number of strangers is exploratory. The reason for setting the highest mark at a relatively large (even large enough to make the real-life situation unrealistic) number of 1000 is we want to make sure that the threshold is reached if there is any. That is to say, apart from those who adopted "Selfish" rationale, all other participants should determine that the amount of strangers were too much to sacrifice even for the sake of their loved ones. Judging from the result, our intended goal was met, and we were able to lower the threshold cap to 100 strangers. But such widespread range also came at a cost. We can't pinpoint the thresholds at a more precise level. In future studies, however, we can narrow the range and thus make smaller intervals between 1-100 if we intend to investigate more thoroughly.

We intended to use the timer as a pressure and to mimic real-life situations where we really need to respond quickly to such dilemmas. We also recorded response time as an extra measure to assess the difficulty for participants to make decisions. Such measure didn't give us a clear and convincing result. The two main reasons are potential confound of presenting sequence and the intervals. Because of the fixed sequence, participants would find it easier in subsequent questions once they anticipated the question patterns, making them use less time to reach their decisions. The wide interval of stranger numbers also made it hard to draw any correlation to test our hypothesis. Fortunately our current results would enable us setting up smaller intervals in the future.

We identified 4 major principles in people's moral reasoning with their

loved ones involved, but it's only at a descriptive level now and we haven't formed a statistical model yet. This is a small step towards a domain that is rarely touched, and we could try to run a factor analysis if we are going to test it. Prior to the experiment we thought that people would always prioritize the benefit (or avoid the loss) of the one they care about first, and should show reliable preference in 1-on-1 situations, but the Recovery principle in real-life situations gives us another perspective. Instead of focusing on preventing the loss, Recovery principle focuses on recovering after the loss. It suggests that it would be easier to recover the loss of someone we care about than some random strangers, so it would also be easier to accept such loss happening to someone we knew well, hence they would bias towards the strangers, contradicting our predictions. Future studies can put more emphasize on such "inverse" trend, to see how it interacts with other major principles like Responsibility, and whether it still holds in other real-life settings.

REFERENCE

Cao, F., Zhang, J., Song, L., Wang, S., Miao, D., & Peng, J. (2017). Framing effect in the trolley problem and footbridge dilemma: Number of saved lives matters. *Psychological Reports, Vol. 120(1),* 88-101

Christensen, J.F. & Gomila, A. (2012) Moral dilemmas in cognitive neuroscience of moral decision-making: A principled review. *Neuroscience and Biobehavioral Reviews, 36,* 1249-1264.

Gold, N., Colman, A.M., & Pulford, B. D. (2014). Cultural differences in responses to real-life and hypothetical trolley problems. *Judgment and Decision Making, Vol.9, 1,* 65-76.

Gold, N., Pulford, B. D., & Colman, A.M. (2013). Your money or your life: comparing judgments in trolley problems involving economic and emotional harms, injury and death. *Economics and Philosophy, 29(02),* 213-233.

Gold, N., Pulford, B. D., & Colman, A.M. (2015). Do as I Say, Don't Do as I Do: Differences in moral judgments do not translate into differences in decisions in real-life trolley problems. *Journal of Economic Psychology, 47,* 50-61.

Hewstone, M., Rubin, M., & Willis, H. (2002). Intergroup Bias. *Annual Review of Psychology, 53,* 575-604.

Miller, J. G., & Bresoff, D. M. (1998). The role of liking in perceptions of the moral responsibility to help: A cultural perspective. *Journal of Experimental Social Psychology, 34,* 443-469.

Patricia O'Neill, & Petrinovich, L. . (1998). A preliminary cross-cultural study of moral intuitions. *Evolution & Human Behavior, 19(6),* 349-367.

Tassy, S., Oullier, O., Mancini, J., & Wicker, B. (2013). Discrepancies between judgment and choice of action in moral dilemmas. *Frontiers in Psychology, Vol.4.*