Journal of Behavioral and Experimental Finance 37 (2023) 100735

Contents lists available at ScienceDirect

Journal of Behavioral and Experimental Finance

journal homepage: www.elsevier.com/locate/jbef

Full length article

Risk-taking for others: An experiment on the role of moral discussion☆

Francesco Feri[a], Caterina Giannetti[b,*], Pietro Guarnieri[b]

[a] Department of Economics, Royal Holloway University of London, United Kingdom
[b] Department of Economics and Management, University of Pisa, Italy

ARTICLE INFO

Article history:
Received 5 October 2021
Received in revised form 30 July 2022
Accepted 1 August 2022
Available online 17 August 2022

JEL classification:
G02
G32
C91
D81

Keywords:
Moral discussion
Mediation analysis
Risky-dictator game

ABSTRACT

We study the effect of *moral discussion* when risk-taking choices entail a negative externality on others. In our experiment, the decision-maker chooses between two risky gambles, one of which entails a better outcome for himself but higher risk for the receiver. In the *Moral Discussion* treatments – before making a choice – decision-makers discuss the consequences of their choice within a group of peers. We also implement a *Reflection* treatment, where participants have to think before making their choice, and a baseline with an immediate decision. Our results show that, after a moral discussion, decision-makers choose more often the less risky gamble compared to the *Reflection*. Moreover, this effect does not depend on the mode of interactions among participants. Through a mediation analysis, we also show that this effect mainly unfolds through a significant modification of the beliefs about the behaviour of their peers.

1. Introduction

One root cause of the financial crisis of 2008 was the excessive risk-taking by many banks and financial agents (e.g. Diamond and Rajan, 2009; De Bruin et al., 2018), i.e. systemic risk. The recognition that agent risk choices can lead to systemic risk has led to active discussions among policy-makers and regulators about financial market reforms, giving rise to prominent ethical issues. Indeed, it is now widely accepted that the financial crisis was not just a crash (e.g. Zingales, 2015; Mian and Sufi, 2017, Group of Thirty 2018) and that failure by banks and bank agents to meet ethical standards played a significant role in it. Thus, "*to what extent do financial agents have a moral duty to limit their contributions to systemic risk?*" (see "*Philosophy of Money and Finance*" in the Stanford Encyclopedia of Philosophy, De Bruin et al., 2018).

On the one hand, it is possible to argue that financial transactions always carry risk and that this is part of the game (as in the famous quote "no guts, no glory", see e.g. Eriksen and Kvaløy, 2017). However, on the other hand, systemic risk can lead to the collapse of the entire financial system by generating severe negative effects on third parties (i.e. negative externalities). Systemic risk has ethical implications and should lead financial agents to commit to caution and social responsibility in order to prevent the spread of social damage (De Bruin et al., 2018; Linarelli, 2017, James, 2017).

The acknowledgement of this moral duty has generated a growing interest in the provision of culture and ethics within the financial sector as a measure to contrast widespread tolerance of dishonest behaviour (Morris and Vines, 2014; Guiso et al., 2015 Klooster and Meyer, 2016; Cohn et al., 2014, 2017; Egan et al., 2019, Suss et al., 2021). Indeed, it has been shown that poor culture leads to greater bank-risk (Kanagaretnam et al., 2019, Suss et al., 2021). Importantly, bank culture and behaviour are now considered key components of financial supervision (De Nederlandsche Bank, 2015; Fernández Muñiz et al., 2018). For example, De Nederlandsche Bank (DNB) is the first supervisory authority in the world to incorporate these aspects into its supervision. Despite this, framing policy debate around these concepts is still viewed as somewhat impractical (Wehinger et al., 2012; De

☆ The authors thank the Einaudi Institute of Economics and Finance, the Institute Europlace de Finance, the Chair of *Individuals and Risk: analysis and market response*, as well as the Chair's partners (Fondation du Risque, Université Paris-Dauphine, ENSAE and Groupama) for financial support. The authors also thank the editor, two anonymous referees, Maria Bigoni, Antonio Cabrales, Fabrizio Ferrero, Miguel Meléndez-Jiméne, Trude Mykleblust, Lorenzo Sacconi, Nicola Salvati, Matthias Sutter as well as participants at the IDEE meetings in Bologna, "Finance and Justice" Conference in Bayereuth, SOCRATES workshop in Pisa, Society for Experimental Finance Conference in Nice, Oikos Young Scholars Finance Academy in Zurich, ESA Conference in Wien, ESA Conference in Berlin, Inaugural Conference of the Experimental Economics Group in Bonn, "Economics and Psychology" Workshop at the Bank of England, for useful comments and suggestions. The authors also thank Veronica Rattini for her excellent work in programming the experiment.

* Corresponding author.
E-mail addresses: francesco.feri@rhul.ac.uk (F. Feri), caterina.giannetti@unipi.it (C. Giannetti).

https://doi.org/10.1016/j.jbef.2022.100735
2214-6350/Crown Copyright © 2022 Published by Elsevier B.V. All rights reserved.

Nederlandsche Bank, 2015, D'Acunto, 2018; Kanagaretnam et al., 2019). Rules-based approaches to financial law and regulation appear to have limited impact. Indeed, in the last few years, the financial services industry has produced several "codes of conduct" albeit with few concrete results,[1] probably as a result of the persistence of a risk-culture and irresponsible behaviour within the banking community (Reynolds and Newell, 2011; Lo, 2015; Young et al., 2012; Cohn et al., 2014; Awrey et al., 2013; Kanagaretnam et al., 2019) and of the incentive schemes that appear ill-suited to resolve agency problems in this sector (Young et al., 2012; Awrey et al., 2013, De Nederlandsche Bank, 2015; Kirchler et al., 2018). Thus, these failures justify the need for new and alternative control instruments, besides incentives and rules, to enhance social responsibility of banks internally and in a way that is self-sustaining.

This paper contributes to the debate with a laboratory study on whether moral discussion favour the emergence of pro-social norms and behaviour in decisions that involve risk-taking for others. To this end, we develop a novel experimental design based on a risky dictator-game in which there is a conflict of interest between a decision-maker and a passive receiver of the consequences of the decision. Indeed, in many banking and financial decisions, the individual remuneration scheme for the delegated person may incentivize risk-taking by exploiting the asymmetry between the risks and returns of the parties involved (Moore and Loewenstein, 2004). The agent is often not exposed to personal losses if a transaction is not successful but could be tempted by the opportunity of personal extra-gains to undertake riskier investments on the shoulder of the client (see Reynolds and Newell, 2011). Likewise, in our experiment the decision-maker may be considered ethically responsible: the difference between the payoffs connected to the two investment options univocally identify the riskier option as the one which is unfair for their counterpart, so that the decision-maker realizes that the decision at stake is to choose between ethical conduct and one which is self-interested. As in Bénabou et al. (2018), the essence of morality concerns actions that may produce "positive externality on others, or avert negative ones" (i.e. a utilitarian definition Gert and Gert, 2017).[2]

We conduct the following treatments: a *Baseline* treatment, in which subjects are asked to choose within a relatively short time; a *Reflection* treatment, in which subjects have to choose after a minimum specified time (longer compared to the baseline treatment); a series of *Moral discussion* treatments (with different communication settings) in which subjects decide after discussing with a peer both the intentions and consequences of their choice on the payoff of both players. In all treatments, however, the final decision is personal and the responsibility of the decision is never shared (as in Cason and Mui, 1997; Bolton et al., 2015, Eijkelenboom et al., 2019).

The comparison between the different treatments allows us to test the hypotheses that a moral discussion with a peer and longer deliberation times make fair choices more likely. As explained in Sections 2 and 4, our main hypothesis about the effectiveness of moral discussion can be easily reconnected to rationalistic moral traditional philosophies, including obviously deontology (Kant 1785) and utilitarianism (Mill 1863) that assert that reasoning fosters moral decisions. This is also in line with Habermas 's recognition (1990) of the fundamental discursive nature of ethics and with his discourse principle. We also reconcile our hypotheses with the recent evidence on the Dual Process Theory (Haidt, 2001).

Our results strongly suggest that moral discussion promote fairer choices, while we do not find convincing evidence for the increase in frequency in the reflection treatment. Moreover, we do not find any systematic differences among alternative types of moral discussions thus suggesting that the mode of peer interaction does not matter. These results are reflected in first and second-order beliefs: compared to *Baseline* and *Reflection* treatments, in *Moral discussion* treatments subjects believe that fair decisions by others are more likely and – at the same time – that others expect from them fairer decisions too. Finally, relying on a mediation analysis (Imai et al., 2010a,c), we find evidence that the effects on decisions are mediated by the descriptive (i.e. empirical) norms, i.e. the expectations about what others will do (first-order beliefs). We do not find evidence that decision-maker's beliefs on what the others expect from them (second-order beliefs) play a role. These results are consistent with the hypothesis that a descriptive norm referred to the reference group of peers drives individual decisions, and that moral discussion help the subject to identify it.[3]

The paper is organized as follows: Section 2 describes the relevant literature, Section 3 presents the experimental design, while Section 4 highlights our hypotheses. Section 5 illustrates and discusses the main results, including both univariate and multivariate analysis of individual choices; Section 6 provides the conclusions of the study.

2. Literature review

This study can be situated in the extensive field of research on social preferences, and more specifically in the subfield that investigates how moral and social norms may promote prosocial behaviour and human cooperation (i.e. the so-called social norms approach, see Hillenbrand and Verrina, 2019, Fehr and Schurtenberger, 2018; Bicchieri, 2006; Krupka and Weber, 2013). More specifically, the review will examine previous research which show that social norms need to be activated to become salient, i.e. to be the benchmark of subjects and affect their behaviour (Fehr and Schurtenberger, 2018; Kallgren et al., 2000; Cialdini et al., 1990).

To the best of our knowledge, few papers explore the effects a moral discussion between peers facing the same dilemma has on the behaviour of the decision-maker. The closest research to our paper is Gunia et al. (2012). The authors investigate the effect of reflection and discussion with a peer using a modified version of the Cheap Talk Sender–Receiver Game in Gneezy (2005).[4] Their results suggest that both contemplation and conversation with an artificial partner led to a higher frequency of "honest" messages compared to the case with an immediate message. However, there are significant differences with our research: we rule out the strategic dimension (i.e. the receiver in our case is entirely passive), and we allow for a genuine discussion among decision-makers (i.e. we do not suggest any type of norm with predetermined messages).

Another related study is Andersen et al. (2018) who compare the immediate choices of participants in a dictator (and cheating) game with those made after a day, in which participants "slept on

[1] A prominent example is the Code of Ethics and Standards of Professional Conduct of the Chartered Financial Analyst Institute.

[2] In other words, in our setting, there is no need to know the risk-preferences of the receiver (e.g. Chakravarty et al., 2011) as it is possible to infer which choice would be the best for them, independently of their attitude towards risk. Importantly, the receiver cannot infer ex-post whether the bad result is due to the risk-seeking behaviour by the decision-maker or to bad luck (moral wiggle room, see Dana et al., 2007).

[3] An alternative explanation is related to guilty-aversion.

[4] They name reflection as *Contemplation*, which is defined as "individually conducted moral reasoning", while discussion with a peer is named *conversation*, defined as "social contemplation".

it". They find that having this additional day – thus longer deliberation time – does not affect the giving (and cheating) decision, further suggesting that any discussion with peers outside the lab environment (as participants may indeed have done) would not change the results either. However, as far as our study is concerned, it is important to notice that the (potential) discussion of the participants in this experiment, unlike our experiment, does not necessarily have a moral focus, and therefore may have not triggered any moral evaluation by the subjects (as suggested by Cialdini et al., 1990, Kallgren et al., 2000). Furthermore they cannot distinguish subjects that only engaged in reflection from those that discussed their decision with peers, neither they can control whether the (potential) discussion occurred with people who were not involved in the experiment. Therefore our experiment represents an advancement of their results as we can disentangle the effects of engagement with reflection from those ensuing from discussion with a peer facing the same decision.

Our research is also closely related to the debate in moral psychology on the Dual Process Theory of moral judgment (see Greene, 2014 and Moore and Tenbrunsel, 2014 for a comprehensive review). This debate centres on the role of decision times and discussion on ethical choices. On the one hand, following the Kantian tradition, moral development theorists argue that moral actions become self-evident through careful deliberation (Kohlberg, 1976; Moore and Tenbrunsel, 2014; Moore and Loewenstein, 2004). On the other hand, the moral intuitionist perspective, as put forward by the seminal contribution by Haidt (2001), asserts that moral decisions are made intuitively, especially when there is a clear social norm (as in Gunia et al., 2012). These two strands of research are often in contrast and difficult to reconcile, especially because the empirical evidence is not always redeeming (e.g. Krajbich et al., 2015), and free from research flaws and design. Recent research, however, suggests that moral decisions are those that have been thought over "just enough" being the relationship between moral choices and decision times curvilinear, with cognitive complexity and personal perspectives playing a key role (Moore and Tenbrunsel, 2014; Frank et al., 2019).

Taken together, the findings of these studies suggest that reasoning on moral dilemmas is positively related to moral actions, although high cognitive complexity or easy-justifiable more dilemmas can disrupt the power of reasoning (e.g. Frank et al., 2019, Bicchieri and Dimant, 2019). This is also in line with Habermas (1990)'s recognition of the fundamental discursive nature of ethics and with his discourse principle: only those norms that meet (or can meet) the approval of all affected participants in the practical discourse can be claimed to be valid.

3. Experimental design

In our experiment a decision maker (type B participant) faces a choice between two lotteries, *Left* and *Right*, whose outcomes determine payments for both the decision maker and a passive receiver (type A participant).

The payoffs and the probabilities of the two lotteries are reported in Table 1.[5] Applying a simple measure of risk of a lottery, as for example, the coefficient of asymmetry (i.e the third moment of a distribution) as well as that proposed by Jia and Dyer (1996), we can rank the two lotteries according to the level of risk they entail and state that choice *Left* entails a higher risk for the passive receiver. A further characteristic of these lotteries is that, for the decision maker, lottery *Left* first-order stochastically dominates lottery *Right*, as well as for the passive receiver, lottery *Right* first-order stochastically dominates lottery *Left*. Therefore, under the assumption of selfishness, the decision makers prefer the riskier lottery *Left*, as just the receivers prefer the safer lottery *Right*. The decision maker, by choosing *Left*, can get a better deal at the cost of a worse lottery for the passive receiver (riskier and first order stochastically dominated). Therefore, the decision-maker faces a trade-off between a self-interested and unfair choice and a fair one, and thus is confronted with an ethical decision in the broad sense used in this paper (see Bénabou et al., 2018, Fehr and Schurtenberger, 2018).

Table 1
Rolling the Dice: Left or Right Choice.

Prob.	*Dice result*	LEFT		RIGHT	
		A	B	A	B
$\frac{1}{6}$	$=1$	6	16	0	6
$\frac{5}{6}$	$\neq 1$	0	6	6	6

If the decision maker chooses "*Right*" and rolls a 1, she gets 6 Euro and the recipient 0 Euro. If the decision maker chooses "*Right*", and rolls a number different from one, she gets 6 Euro and the recipient 6 Euro.
On the other hand, if she chooses "*Left*" and rolls a 1, she gets 16 Euro and the recipient 6 Euro. If she chooses "Left" and rolls a number different from one she gets 6 Euro and the recipient 0 Euro.

In addition, since the decision concerns lotteries and not certain payoffs (with the same incentive structure), it is difficult for the passive receiver to infer from the realization of the payoffs which option the decision maker has chosen. This gives the decision maker the possibility to hide the action and stronger incentives for a selfish choice (moral wiggle room, Dana et al., 2007).

Upon their arrival, participants were randomly assigned to one of two rooms. Each participant in each room knew they had been randomly and anonymously paired to another subject in the other room. We used the strategy method whereby each participant in each room had to decide between *Left* and *Right* without knowing their role. Participants also knew that participants in the other room were facing the same decision under the same conditions. Only at the very end of the experiment were participants in one room randomly assigned the role of decision maker (i.e. type B). All the participants in the other room were assigned the role of receiver respectively (i.e. type A).[6] Therefore, the decision of each B-player in the selected room determined the payoff of the paired A-player. See the English translation of the instructions at the end of this paper.

After they made their decision, and before knowing their role in the experiment, participants were also asked about their beliefs.[7] More specifically, participants were asked to specify their beliefs concerning: (1) the percentage of subjects choosing *Left* in the other room (first-order belief); (2) the average response to the previous question in the other room (second-order belief). Payments to this phase were determined by a lottery selecting only one of the questions and by rewarding those who had correctly answered to the selected question within a range of 10% tolerance.[8]

[5] Lotteries are implemented by rolling a six-face dice.

[6] Roles were assigned by drawing a card from a card deck, the first room to pick a red card was assigned the B status.

[7] We choose to ask about beliefs before finding out their role to rule out any possible interference on the beliefs.

[8] In addition we asked the percentage of subjects choosing left in their same room ("first-order belief - same room")

More precisely we asked the following questions:

1. Out of 10 participants, how many participants do you believe have played left in this room?

2. Out of 10 participants, how many participants do you believe have played left in the other room?

3. Questions 2 was asked to the participants in the other room. What do you think is the average answer to that question?

Table 2
Treatment Overview.

Treatment	Time		Decision
Baseline (BT)	4 minutes[a]		Left or Right choice
Reflection (RT)	4 minutes[a]	4 min alone	Left or Right choice
Moral Discussion (MDT)			
Moral 2	4 minutes[a]	4 min discussion in pairs	Left or Right choice
Moral 3	4 minutes[a]	4 min discussion in a group of three	Left or Right choice
Moral Chat	4 minutes[a]	4 min discussion in pair via a chat	Left or Right choice

[a]Clarifications questions to the experimenter are allowed.

In order to identify the effects of deliberation and moral discussion, we run several treatments differing in time and conditions of decision-making. More specifically, in *Baseline Treatment* (BT), participants were given four minutes in order to let them fully understand the instructions and ask the experimenter clarification questions ["*You have now 4 min to re-read carefully these instructions. During this time, if you have questions please rise your hand and we will personally answer you.*"]. In the *Reflection Treatment* (RT), participants were given 4 min to understand the instructions, and then 4 additional minutes to think individually and in complete silence [*"After the first 4 min, you have other 4 min to think alone and in complete silence.*"]. In the *Moral Discussion Treatments* (MDT), subjects were also assigned to a group of participants in the same room, i.e., their neighbour(s), to talk about the consequences of their decisions for 4 min (after the first 4 min to understand the instructions), as well as their personal intentions concerning the decision at stake [*"After the first 4 min, you can talk with the participant seated next to you for other 4 min*" about *"the consequences that your decision will have on player-A and player-B" (but only with him/her). You can also discuss your intentions with him/her regarding the decision whether to play right or left. At the end of the discussion you will have to make your choice, right or left. Notice that this final choice will be private and no one, including your discussion mate, can see it*"].[9] Therefore, in MDT treatments participants not only had more time as in RT to think about their choice but they also need to engage in a moral discussion with a peer. In particular, in *Moral 2* participants were grouped in pairs for face-to-face interaction, while in *Moral Chat* participants were grouped in pairs but had to converse via a chat in order to check for any differences in communication mode and to track the content of their conversation. In *Moral 3*, participants were in a group of three for face-to-face interaction in order to check for any differences due to group size. In all MDT treatments, after the 4 min of discussion, participants were asked to turn back to their screen and make their choice individually. See Table 2 for an overview.[10]

Table 3
Social Preference Choice.

		X	Y
Line 1	*You*	2	2
	Your partner	2	1
Line 2	*You*	2	3
	Your partner	2	1
Line 3	*You*	2	2
	Your partner	2	4
Line 4	*You*	2	3
	Your partner	2	5

Numbers represent Euros.

The comparison of RT with BT measures the effect of deliberation on the choice of the decision-maker, and the comparison between MDT and RT measures the effect of the moral discussion with a peer (or with peers). Finally, the comparison between MDT and BT gives us a measure of the joint effect of deliberation and moral discussion.

In addition, we control in all treatments for the pre-existence of social preferences through a set of lotteries as discussed in Bartling et al. (2009).[11] Specifically, at the beginning of the experiment, each subject was exposed to 4 decisions in which she had to choose how to allocate payoffs between herself and another subject, randomly and anonymously paired to her in the same room. Everyone had to choose between allocation X and Y (see Table 3). The results of these lotteries were given only at the very end of the experiment.[12] A questionnaire with a short version of the big-five questions (John et al., 1991, 2008), and relevant personal information (sex, age, years of university attendance) concluded the experiment.

The experiment took place at the "Laboratorio di Economia Sperimentale" of the University of Pisa on January and May 2017. We conducted 16 sessions, each involving either 28, 24 or 20 participants, for a total of 412 participants invited from a pool of more than 1500 registered students from every department of Pisa University. None of the students could take part in more than one session. The average pay was 10,90 €, including the show-up fee of 5 €. In total, we ran 2 sessions of the *Baseline* treatment (BT), 2 sessions of the *Reflection* treatment (RT), and 12 sessions with different types of the *Moral Discussion* treatments (MDT). In particular, we conducted 3 sessions of the moral discussion between two people (*Moral 2),* 4 sessions of the moral-discussion

[9] It is important to notice that moral discussion may involve a risk of *experimenter's demand effect*, which is common to experimental studies investigating the effect of normativity and moral framing on decision making. In our experiment, by organizing a"moral discussion" among participants, we make the group conversation internal to the game, thus making the normative definition of what is appropriate emerge from the interaction among subjects and making it hard for subject to understand what we expect them to play. This way, not only can we avoid the problem of a normative (demand) effect, but we can also analyse factors determining (from within) the *formation* of the norm (see Appendix A4). Finally, the high monetary incentives are difficult to forego just to make the experimenters happier.

[10] In all treatments the experimenter had read the instructions aloud at the beginning.

[11] On the importance of controlling for social preference in this context see Krajbich et al. (2015).

[12] At the end of the experiment, only one pair in each room was selected, then one decision line was randomly selected for payment.

composed of two persons communicating *via* chat (*Moral Chat)*, and 5 sessions of the moral-discussion composed of three persons (*Moral 3).* An overview of participants' characteristics is available in Table (A1) in the Appendix.

To increase the power of our analysis we also conducted 7 additional sessions in the same laboratory between 24 February 2022 and 3 March 2022 for a total of 112 participants (average pay was 10,70 €). We ran 2 additional sessions for RT, and 5 additional sessions for MDT, in particular for the *Moral Chat*. We opted to increase the sample size only for these two treatments for two specific reasons. First of all, among all comparisons, this one reflects our main hypothesis and the scope of our research. In addition, due to COVID restrictions, the other Moral treatments (*Moral 2* and *Moral 3*) could not be conducted in the same way as in the past (larger physical distance + facial mask). To allow for a cleaner comparisons of all results, we summarize the results of these additional sessions in Section 5.1. For an overview of the characteristics of these new group of participants see Table (A.2) in the Appendix.[13]

4. Hypotheses

Henceforth, we denote the *Left* (*Right*) lottery as the *Unfair* (*Fair*) lottery, consequently, the decision to choose the Unfair (Fair) lottery is denoted as Unfair (Fair) decision or choice.

In our experiment, the choice the decision-maker faces can be assimilated (as in Gunia et al., 2012) into a "*right-wrong decision*" which is a specific type of moral decision between an intrinsically ethical course of action, i.e. an action which reflects a moral value (e.g. honesty, being fair), and unethical behaviour, i.e. the possibility to deviate from the normative moral value for self-interested gain (e.g. self-interested lying, risk on others). As discussed in Section 2, in this case, moral decisions may require an effortful cognitive process. In other words, serious thought for a period of time gives individuals time to actively consider values associated with the less tempting choice.

Recently evidence on the Dual Process Theory of moral judgement also suggests that there is a "right amount of time" for reasoning on moral choices, whereby the relationship between moral choices and decision times is curvilinear. In other words, reasoning on moral dilemmas drives moral actions, but it can backfire in the case of high cognitive complexity or easy-justifiable moral dilemmas (e.g. Frank et al., 2019).

Thus, we expect longer deliberation times to favour the emergence of prosocial behaviour. Our hypothesis rests on the considerations that reflection enhances the *cognitive awareness of relevant moral values* and *the identification of the consequences of the decision,* and on the hypothesis that the relevant moral norm is "*do not exploit others to get more benefits*" (see also Bénabou et al., 2018). Then, through deliberation, this moral norm becomes salient to decision-makers, compensating for their consolidated selfish attitude.[14] Thus we can state our first prediction:

Prediction 1. Under the assumption of a common moral norm supporting prosocial behaviour, the frequency of *unfair* decisions in RT will be lower than in BT.

We also expect this effect to be strengthened by moral discussion. As in the concept of "discourse ethics" in Habermas (1990), the solution of ethical problems and the identification of substantial ethical norms emerge from moral discussions in which participants, through interactive argumentative procedures, reach a consensus. In addition, in order to observe a change in behaviour, norms need to be activated and become the focus of the subjects (Kallgren et al., 2000; Fehr and Schurtenberger, 2018; Cialdini et al., 1990). As a result, the frequency of unfair decisions is further reduced after a discussion with peers. We can state our second prediction:

Prediction 2. The frequency of *unfair* decisions in MDT will be lower than in RT.[15]

We note that the identification of the consequences of the decision and of the relevant moral norm may affect the beliefs about the behaviour of others. In detail, we expect that the effects described in the previous predictions to be reflected in what subjects expect that others will choose and what subjects expect that the passive receivers of the consequences of the decision expect from the decision makers. So we can state our third prediction.

Prediction 3. The frequency of *unfair* decisions in the first and second order beliefs in MDT will be lower than in RT. Furthermore, these frequencies will be lower in RT than in BT.

5. Results

In Table 4 we list the relative frequency of unfair choices by treatment, the number of independent observations per treatment (i.e. one observation per individual in BT and RT, and one observation per pair or threesome in MDT), as well as treatment comparisons. Furthermore, we list the frequency of unfair choices for each type of MDT and the relative differences between RT and BT. In addition to two-sided t-tests, we report *p-values* associated to one-sided *t-test* when our prediction has a precise sign direction (i.e. the alternative hypothesis is of the type $H_1 : \mu_{MMT} > \mu_{RT} > \mu_{BT}$).[16] To check the robustness of our results, we additionally report *p-values* associated to one-sided *permutation* tests (i.e. 1000 data shuffling), as well as confidence intervals.[17] In Section 5.1 we also provide additional evidence derived from new experimental sessions conducted in a different point in time to increase our statistical power, along with equivalence tests (Hoenig and Heisey, 2001, Lakens, 2017).

In line with our hypothesis that moral discussion favour more ethical (pro-socially responsible) choices, we observe in Table 4 that participants in the MDT showed the lowest percentage of unfair decisions (26%), while subjects who were also allowed to reflect by themselves in the RT opted for unfair decisions less often (35%) than subjects in the BT (50%). These values are in line with our hypotheses that longer deliberation times and moral discussions have a positive effect on ethical choices. However,

[13] Due to COVID restrictions, the sessions were smaller and limited to either 8, 12, 16, 24 participants.

[14] Furthermore, the above hypothesis casts a new light on decisions made in groups. If it is true that groups often display a higher strategic capability (Kocher and Sutter, 2005), sustaining also unethical decisions such as deception (Sutter, 2009), it is also possible that group discussion may also promote prosocial behaviour, based on what is the relevant current social norm. In our framework, deliberation and discussion play an active role in re-framing the decision, so that the decision maker, at the end of the process, acquires greater awareness of the consequences of their decision and of the value of the decision itself, and can adhere to previously unacknowledged moral norms.

[15] From a psychological perspective, the hypothesis on the effect of moral discussions and deliberation are in line with Gunia et al. (2012).

[16] In the following, we always rely on *t-test* differences, and therefore in the Normal approximation for a Binomial. The Normal distribution is a good approximation to the binomial when n is sufficiency large and p is not too close to 0 or 1. If p is near 0.5, the approximation can be good for n much less than 20. To be conservative, the normal distribution is of a good use as an approximation to the binomial when $np > 5$ and $n(1 - p) > 5$.

[17] Permutation tests are similar to a placebo test. If the null hypothesis of no treatment effect is true, changing the exposure to the treatment would have no effect on the outcome. Therefore, by randomly shuffling the exposures we can derive the sampling distribution of the test statistic without imposing any parametric distribution on the outcome. If the null hypothesis is true the shuffled data sets should look like the real data.

Table 4
Relative frequencies of unfair decisions by treatment.

Treatment	Indep Obs	Unfair Choices		Diff	*Two-sided p-value*	*One-sided p-value*	*Permutation p-value*	*95% Conf. Interval*
Baseline (BT)	56	0.50	**Moral vs Baseline**	−0.24	0.000	0.000	0.000	[−0.35 −0.12]
Reflection (RT)	56	0.36	**Reflection vs Baseline**	−0.14	0.126	0.063	0.046	[−0.33 0.04]
Moral Discussion (MDT)	130	0.26	**Moral vs Reflection**	−0.10	0.102	0.051	0.050	[−0.21 0.02]
Moral 2	42	0.27	*Moral 2 vs Baseline*	−0.23	0.014	0.007	0.005	[−0.41 0.04]
			Moral 2 vs Reflection	−0.09	0.276	0.173	0.137	[0.26 −0.09]
Moral 3	40	0.31	*Moral 3 vs Baseline*	−0.19	0.000	0.016	0.020	[−0.41 −0.04]
			Moral 3 vs Reflection	−0.05	0.568	0.284	0.297	[−0.22 −0.12]
Moral Chat	48	0.21	*Moral Chat vs Baseline*	−0.29	0.000	0.000	0.001	[−0.46 −0.13]
			Moral Chat vs Reflection	−0.15	0.064	0.032	0.021	[−0.31 0.01]
			Moral 2 vs Moral Chat	−0.06	0.334		0.134	[−0.20 0.07]
			Moral 2 vs Moral 3	−0.04	0.628		0.693	[−0.11 0.18]
			Moral Chat vs Moral 3	−0.10	0.108		0.054	[−0.22 0.02]

The permutation tests *p*-value are computed by counting how many permuted mean-differences out of 1000 permutations are larger than the one we observed in our actual data. The number of independent observations per treatment is obtained by keeping for each group (i.e. pair or triad) one single observation, i.e. the mean of unfair choices in the group. There are therefore, 42 observations (i.e. pairs) in Moral 2, 48 observations (i.e. pairs) in Moral Chat, 40 observations (i.e. triad) in Moral 3.

these effects (which account for correlation at group level) are not significant from a statistical point of view. Specifically, the difference between RT and BT (i.e. −0.14, relative risk = 0.72) and between MDT and RT (i.e −0.10, relative risk = 0.722) are not statistically significant if a *two-sided test* is considered (in both cases p-values> 0.10). The results are statistically significant only when comparing MDT and BT (i.e. −0.24, relative risk = 0.52). The statistical significance of the results, however, is achieved (with p-values below 0.050) when looking at one-sided test and 1000 permutations of the exposure to treatment (which do not rely on any statistical distributional assumptions). This evidence can be summarized in the following statement:

> Result 1: A period of deliberation before taking the decision induces a higher frequency of fair choices. If this period is replaced by a moral discussion with a peer, we observe a further increase in the frequency of fair choices. However, the results are not statistically significant if two-sided tests are used.

Different from Schram and Charness (2015), the individual's choice is never observable by others in our design. In accordance with their results, our result suggests that a moral discussion can allow the emergence of a norm ("a shared understanding about what one ought to do in a specific situation") and may be a more powerful device to affect moral reasoning than receiving a simple advice from external observers.[18]

Furthermore, within the MDT treatments, the largest effect is observed when participants discussed the decision via a chat function, i.e −0.15 (p-value = 0.064, relative risk = 0.58). However, there are no significant differences in unfair choices across modes of interaction and across group size among MDT. It thus seems that moral discussions may help to identify social norms and discussing them appears to be immune to its concrete mode of discussion. Therefore, we can state the following result:

> Result 2: There are no significant and systematic differences among MDT treatments.

Table 5 reports the average of elicited beliefs by treatment, i.e. the expected relative frequency of unfair decisions. The upper panel reports first-order beliefs, i.e. the subjective probability that others will take an unfair decision; the bottom panel reports second-order beliefs, i.e. the beliefs on the other subjects first order beliefs. Differences across treatments are also reported, as well as the beliefs for each of the MDT together with their differences with RT.

We observe striking differences across treatments: in MDT subjects have first order beliefs which are significantly lower compared to the other treatments, i.e. their expectation that others will choose the unfair choice is significantly lower in MDT with respect to the other treatments (see upper panel of Table 5). Indeed, the share of people believed to play left in the other room (first-order belief) decreases from 56% of the BT, to 55% of the RT, to 44% of the MDT.[19] Therefore, we can assume that subjects believe a that moral discussion is effective in order to influence the decisions of others in the direction of more fair choices. Furthermore, these results are also consistent with a general feeling that reflection is not a good device to induce more fair decisions.

These results are reflected in the second-order beliefs (reported in the bottom panel of Table 5). We observe that second-order beliefs in MDT are significantly lower respect to those in both RT and BT (with a difference, respectively, of 13% and 11%). Again we observe no statistically significant difference across RT and BT.

In general, it seems that a moral discussion has a strong effect in shifting participants expectations concerning the behaviour of other subjects, from a more self-interested vision to a more pro-social and ethical. Thus, we can state our third result:

> Result 3: After a moral discussion with a peer, subjects have higher expectations that others will adopt fair decisions and believe that others hold higher expectations from them as well. A period of deliberation has no effect on beliefs.

Comparing decisions (Table 4) with the first order beliefs (upper panel of Table 5) we also note that the latter is systematically higher than the share of unfair decisions. It means that there are subjects that opted for the fair choice even when believing that a non negligible share of people in the other room were choosing the unfair one. Thus, this evidence tentatively suggests that there are subjects who think they are "morally superior" to others and prefer to stick to a norm of behaviour.

[18] In line with Schram and Charness (2015), who find that females are more likely to follow advices, we also observe a larger effect of moral discussions with respect to RT when the group only comprises females, but we consider these results as a simple suggestive evidence since the statistical power of this analysis is very low. See also Section 5.1. Results are available upon request.

[19] We find very similar results for first-order beliefs in the same room. See Table (A.4) in Appendix A.

Table 5
Beliefs about Unfair decisions by treatment (relative frequencies).

	First-order beliefs							
Treatment	Indep Obs	Unfair Choices		Diff	*Two-sided p-value*	*One-sided p-value*	*Permutation p-value*	*95% Conf. Interval*
Baseline (BT)	56	0.56	**Moral vs Baseline**	−0.13	0.000	0.000	0.000	[−0.18 −0.07]
Reflection (RT)	56	0.55	**Reflection vs Baseline**	−0.01	0.744	0.372	0.072	[−0.10 0.07]
Moral Discussion (MDT)	130	0.44	**Moral vs Reflection**	−0.11	0.000	0.000	0.000	[−0.17 −0.05]
Moral 2	42	0.45	*Moral 2 vs Baseline*	−0.11	0.014	0.007	0.001	[−0.19 −0.02]
			Moral 2 vs Reflection	−0.10	0.036	0.018	0.027	[−0.18 −0.01]
Moral 3	40	0.43	*Moral 3 vs Baseline*	−0.13	0.002	0.001	0.000	[−0.21 −0.04]
			Moral 3 vs Reflection	−0.12	0.008	0.004	0.005	[−0.20 −0.03]
Moral Chat	48	0.42	*Moral Chat vs Baseline*	−0.14	0.000	0.000	0.000	[−0.22 −0.06]
			Moral Chat vs Reflection	−0.13	0.002	0.001	0.001	[−0.21 −0.05]

	Second-order beliefs							
Treatment	Indep Obs	Unfair Choices		Diff	*Two-sided p-value*	*One-sided p-value*	*Permutation p-value*	*95% Conf. Interval*
Baseline (RT)	56	0.57	**Moral vs Baseline**	−0.11	0.000	0.000	0.000	[−0.16 −0.05]
Reflection (BT)	56	0.59	**Reflection vs Baseline**	−0.02	0.578	0.289	0.334	[−0.05 0.10]
Moral Discussion (MDT)	130	0.46	**Moral vs Reflection**	−0.13	0.000	0.000	0.000	[−0.18 −0.07]
Moral 2	42	0.48	*Moral 2 vs Baseline*	−0.09	0.018	0.009	0.005	[−0.16 −0.02]
			Moral 2 vs Reflection	−0.11	0.004	0.002	0.015	[−0.19 −0.03]
Moral 3	40	0.44	*Moral 3 vs Baseline*	−0.13	0.002	0.001	0.006	[−0.20 −0.04]
			Moral 3 vs Reflection	−0.15	0.000	0.000	0.000	[−0.22 −0.06]
Moral Chat	48	0.45	*Moral Chat vs Baseline*	−0.12	0.002	0.001	0.000	[−0.18 −0.04]
			Moral Chat vs Reflection	−0.14	0.000	0.000	0.001	[−0.21 −0.06]

The permutation tests *p*-value are computed by counting how many permuted mean-differences out of 1000 permutations are larger than the one we observed in our actual data. The number of independent observations per treatment is obtained by keeping for each group (i.e. pair or triad) one single , i.e. the mean of unfair choices in the group. There are therefore, 42 observations (i.e. pairs) in Moral 2, 48 observations (i.e. pairs) in Moral Chat, 40 observations (i.e. triad) in Moral 3.

Finally, in order to rule out that peer discussions simply help subjects gain a better understanding of the instructions and to provide some evidence of the moral content of the conversations, we performed a text analysis of the chats from one of the MDT (i.e. *Moral Chat* with 48 chats). In particular, we are able to classify three types of messages according to their moral content: Egoistic, Utilitarian and Deontological (all details and definitions are reported in Appendix A.4).[20] We found that almost all chats contain one or more of these types of messages. Therefore, it appears that participants responded to the task by participating in explicit moral reasoning. Thus we can state that the effects of MDT is genuine and that it is not caused by a different use of the discussion.

From the above results, a question about what is causing the increased prosocial behaviour observed in MDT arises: is it a direct effect of the moral discussion? Or is it mediated by beliefs? In particular, if the effect is mediated by the beliefs, it might act through two different channels. The first channel works through first-order beliefs, as these beliefs reflect the perceived descriptive norm that bring people to conform their behaviour to it (e.g. Bicchieri, 2006). The second channel also unfolds through second-order beliefs. Individuals are aware that other people are also similarly conscious of the descriptive norm and expect them to conform to it. This hypothesis is also consistent with guilt aversion theory (Battigalli and Dufwenberg, 2007), i.e. the effect of moral discussion over risk-taking for others depends on decision-makers trying to satisfy what they believe the others expect from them. These questions motivate Section 5.2 devoted to a mediation analysis.

5.1. Additional evidence and power analysis

Before conducting the mediation analysis, we check the robustness of our results. Indeed, an ex-post calculation of the minimum detectable size suggests that we are able to detect a difference in treatment choices of about 21%, with 80% power, considering 50% as a baseline proportion (i.e. the random choice) and a sample size as in RT and MDT. The difference increases to about 25% for a sample size as in BT and RT. By looking at Table 4, we notice that some estimated effects are around 10% (in particular the one of our main interest *Moral* vs *RT*).

Thus, to check the robustness of our results, and to increase the power of our analysis, as stated above (see Section 3), we decided to replicate the two treatments of our main interest, that is RT and *Moral Chat* (within the MDT),[21] and to additionally conduct interval hypothesis testing (equivalence and minimum effect tests). Equivalence tests are a specific implementation of interval hypothesis test, where instead of testing against a null hypothesis of no effect (e.g., an effect size of 0), an effect is tested against a null hypothesis that represents a range of non-zero effect size. As with any hypothesis test, we can reject the equivalence (between treatments) whenever the confidence interval around the observed effect is within pre-determined bounds.[22] Moreover, if the researcher has specified a smallest effect size of interest and, as in our case, is specifically interested in testing whether the effect in the population is larger than this smallest effect of interest, a minimum effect test can be performed. In that case, the estimated confidence interval should be fall completely beyond the smallest effect size of interest.

The results including the additional evidence are reported in Table 6. The first thing to notice is that the share of unfair choices in *Moral Chat* is substantially identical to the one obtained in the 2017 (i.e. 23%), while the share of unfair choices is higher

[20] For a formal discussion about the impossibility to distinguish directly from choices between utilitarian and deontological (Kantian) types see Bénabou et al. (2018).

[21] As stated above, the choice to replicate only Moral Chat is due to Covid restrictions which make impossible to exactly reproduce the *Moral 2* and *Moral 3*.

[22] In other words, the researcher specifies an upper Δ_U and lower Δ_L equivalence bound based, e.g., on the smallest effect size of interest (Lakens, 2017). Then two composite null hypotheses are tested: $H_{01}\Delta < -\Delta_L$ and $H_{02} > \Delta_U$. If both one-sided tests can be statistically rejected, we can conclude that $\Delta_L < \Delta < \Delta_U$, i.e. the observed effect is small enough to be considered equivalent. See Lakens, 2017 also for a discussion about the choice of the bounds.

Table 6
Relative frequencies of unfair decisions by treatment (additional sessions).

Treatment	Indep	Additional	Unfair		Diff	*Two-sided*	*95% Conf. Interval*
	Obs	Indep Obs	Choices			*p-value*	
Baseline (BT)	56	0	0.50	**Moral vs Baseline**	−0.24	0.000	[−0.36 −0.13]
Reflection (RT)	100	44	0.49	**Reflection vs Baseline**	−0.01	0.453	[−0.16 0.18]
Moral Discussion Total (MDT)	164	34	0.26	**Moral vs Reflection**	−0.23	0.000	[−0.33 −0.13]
Moral Chat	82	34	0.23	*Moral Chat vs Baseline*	−0.27	0.000	[−0.41 −0.134]
				Moral Chat vs Reflection	−0.26	0.000	[−0.39 −0.14]

in RT (i.e. 49%), and very much close to that one observed in BT (for which we did not collect further evidence). Therefore, we avoid any strong conclusion for this latter comparison, and leave it out for future research. One possible explanation that deserve further analysis is the possibility of repeating the reflection treatment in which individuals are explicitly told to reflect on the consequences of their choice. Indeed, as it is in the current experimental set-up, we did not induce any type of moral reasoning, and individuals by themselves appear not able to identify any relevant norm of behaviour. Therefore, slowing decision times may not be enough to promote fairer choices (our prediction 1).

On the contrary, this additional evidence reinforces what we have already observed in 2017 about the efficacy of *Moral Discussion* - with respect to *Reflection* - in helping individuals to identify the relevant norms at play, thereby opting for a fairer choice (our prediction 2). In this case, the difference between RT and *Moral Chat* increases to 26% (p-value = 0.000, relative risk = 0.49), while the difference between RT and MDT increases to 23% (p-value = 0.000, relative risk = 0.53).

As stated above, to further validate our evidence, as suggested by Hoenig and Heisey, 2001 (see also Lakens, 2017), we also report confidence intervals and compute equivalence/non-inferiority tests. The latter aim to capture the maximum level of difference in treatments which is supported by our data. In other words, the term "equivalent" is not in the strict sense, but rather it means that the effects of the two treatments are close enough so that one cannot be considered superior or inferior to the other. In our case, by choosing a lower bound for the minimum effect of −30% and an upper bound of −10%, we can reject the hypothesis of equivalence between our treatments, as the estimated confidence interval [−0.33 −0.10] is outside these bounds.[23] Moreover, as the upper limit of the confidence interval lies below the upper limits of −10%, we can also assert that *Moral Discussion* is not superior (in terms of unfair choices) compared to RT. If we consider the confidence interval of *Moral chat* [−0.39 −0.14] the confidence interval is even larger.

To complete the analysis we also report in Table 7 the average values for the beliefs. In line with the evidence above, thus, we consistently observe a statistical significant reduction of about 10% in both first and second-order beliefs after a moral discussion, while no significant effect is observed compared to the baseline after reflection.

5.2. Mediation analysis

Relying on the original sample, in this section we investigate the determinants of the unfair choice through a mediation analysis. The goal is to determine causal mechanisms by examining the roles of intermediate variables that lie in the causal path between the treatment and outcome variables (Imai et al., 2010a). We aim to quantify how much of our treatment variable (i.e. Moral Discussion) is transmitted to choice by our mediating variable, i.e. beliefs.[24] We disentangle two effects that our treatment variable has on the probability of choosing the *unfair* lottery: one direct effect of the treatment (in the following ADE); and one indirect effect that affects the choice through a modification of the mediator, i.e. participants' beliefs (in the following ACME). See Appendix B for a detailed overview of this methodology. To conduct the mediation analysis, we rely on a linear simultaneous equations model (SEM) by estimating the following system:[25]

$$\begin{cases} FOB &= \beta_1 MMT + \theta_1 BT + \eta_{11} Prosocial \\ & \quad + \eta_{12} Envy + \eta_{13} Envy \cdot Prosocial + \varepsilon_1 \\ SOB &= \beta_2 MMT + \theta_2 BT + \eta_{21} Prosocial \\ & \quad + \eta_{22} Envy + \eta_{23} Envy \cdot Prosocial + \varepsilon_2 \\ Unfair &= \gamma_1 FOB + \gamma_2 SOB + + \eta_{31} Prosocial + \eta_{32} Envy \\ & \quad + \eta_{33} Envy \cdot Prosocial + \beta_3 MMT + \theta_3 BT + \varepsilon_3 \end{cases} \tag{1}$$

where FOB and SOB denote, respectively, first and second-order beliefs, *Unfair* is a variable taking value 1 (0) if the unfair (fair) lottery is chosen. *Envy* and *Prosocial* denote, respectively, the dummy variables for envy and prosocial as described and reported in Table 8, which were derived from the elicitation task in Table 3. Finally *MDT* and *BT* are dummy variables equal to 1 for observations in MDT and BT respectively, and zero otherwise. The baseline category is represented by RT, so that we can have a direct statistical test of the main treatments comparisons (i.e. MDT vs RT, and BT vs RT).

After the estimation of the model, the product-of-coefficients method (i.e. "Barron-Kenny procedure") yields an estimate of the mediation effects by multiplying the relevant coefficients of each equation (Imai et al., 2010c). For example, the direct effect of MDT on choosing *unfair* is $\hat{\beta}_3$, while $\hat{\gamma}_1\hat{\beta}_1$ and $\hat{\beta}_2\hat{\gamma}_2$ can be interpreted as valid estimates of the causal mediation effects that unfold through first and second order beliefs. Similarly, $\hat{\theta}_3$ captures the direct effect of BT, while the mediation effects are captured by $\hat{\theta}_1\hat{\gamma}_1$ and $\hat{\theta}_2\hat{\gamma}_2$. In the same way, one can also compute the indirect effects of social preferences that unfold through beliefs (for example for variable Prosocial indirect effects are $\hat{\gamma}_1\hat{\eta}_{11}$ and $\hat{\gamma}_2\hat{\eta}_{21}$).

Table 9 reports the indirect and direct effects on the probability of *unfair* choice for both our mediators, i.e. first and second order beliefs.[26] The full estimates of the model equations are reported in Table (A.5) in the Appendix A, while a sensitivity analysis, reported in Appendix B.1, show that results are robust to some fundamental but technical assumptions, necessary to estimate the model. As Table 7 highlights, there are

[23] As it appears clear, changing the bounds allow to test for different level of equivalence. Whenever the estimated confidence interval lies within the bounds, it is no possible to reject that the effect is small enough to zero and practically equivalent.

[24] A mediator variable can be thought of as a post-treatment variable that occurs before the outcome is realized (Imai et al., 2010c).

[25] We assume linearity but results are robust if we rely on a non-linear model (e.g. a probit) for the outcome variable *Unfair*.

[26] Estimations are based on non-parametric bootstrap algorithm proposed by Imai et al. (2010a), which returns point estimates essentially identical the product of coefficients.

Table 7
Beliefs about Unfair decisions by treatment (additional sessions).

	First-order beliefs						
Treatment	Indep Obs	Additional Indep Obs	Unfair Choices		Diff	*Two-sided p-value*	*95% Conf. Interval*
Baseline (BT)	56	0	0.56	**Moral vs Baseline**	−0.13	0.000	[−0.18 −0.08]
Reflection (RT)	100	44	0.55	**Reflection vs Baseline**	0	0.517	[−0-08 −0.08]
Moral Discussion Total (MDT)	164	34	0.43	**Moral vs Reflection**	−0.13	0.000	[−0-18 −0.08]
Moral Chat	82	34	0.41	*Moral Chat vs Baseline*	−0.15	0.000	[−0.21 −0.09]
				Moral Chat vs Reflection	−0.15	0.000	[−0.21 −0.09]

	Second-order beliefs						*95% Conf. Interval*
Treatment	Indep Obs	Additional Indep Obs	Unfair Choices		Diff	*Two-sided p-value*	
Baseline (BT)	56	0	0.56	Moral vs Baseline	−0.12	0.000	[−0.17 −0.07]
Reflection (RT)	100	44	0.58	Reflection vs Baseline	0.02	0.572	[−0.08 −0.04]
Moral Discussion Total (MDT)	164	34	0.44	Moral vs Reflection	−0.14	0.000	[−0.19 −0.10]
Moral Chat	82	34	0.43	*Moral Chat vs Baseline*	−0.13	0.000	[−0.20 −0.08]
				Moral Chat vs Reflection	−0.16	0.000	[−0.21 −0.10]

Table 8
Distribution of social preferences.

	Envy	Not Envy	Total
Prosocial	202	164	366
	(55%) [84%]	(45%) [95%]	(100%)
Not Prosocial	38	8	46
	(83%) [16%]	(17%) [5%]	(100%)
Total	240	172	412
	[100%]	[100%]	
(% over row), [%over column]			

Prosocial is a dummy variable equal to 1 if an individual chose X either in Line 1 or Line 2 in Table 3.
Envy is a dummy variable equal to 1 if an individual chose X either in Line 3 or Line 4 in Table 3.

no significant direct effects(as the confidence interval for the ADE coefficients always contains the zero). However, the effect is significant when MDT are compared to BT. This suggests that interaction between reflection and moral discussion has a significant direct effect. Moreover, MDT have a statistically significant indirect effect, compared both to RT and BT, that unfold through a modification of first order beliefs (the confidence interval for the indirect coefficient – ACME – does not contains the zero). This effect is also economically significant since it implies a reduction in the probability of choosing the *unfair* lottery of about 8% (starting from 50% in BT).[27] Consistently with the evidences in Bicchieri and Xiao (2009) this result suggests that empirical/descriptive norms (through first order beliefs) play an important role, while guilt aversion (through second order beliefs) is not relevant in this kind of decision. Finally, we observe that our measure of social preference variables have a direct negative effect on the probability of choosing the *unfair* lottery, while the mediated effects through the beliefs (indirect effect) is not significant. While the observation that the Prosocial and Envy variables may at first sight appear a bit odd, this result basically reflects the degree to which subjects in the experiment are willing to follow a "moral norm" (i.e. *fair choice*) rather than directly their own general preferences over payoff distributions (Kimbrough and Vostroknutov, 2016). In unreported regressions, we repeat the analysis removing pro-social variables, and by replacing prosocial preferences with big-five traits. Results are robust in both cases, though we do not find any significant effects of big-five traits. Thus, we can state our fourth result:

> Result 4: Only the interaction between discussion and deliberation has a significant direct effect on increasing the probability of prosocial choices. There is a significant indirect effect of MDT that unfold through first-order beliefs.

6. Conclusions

In this paper, we introduce a novel experimental design to study a situation in which there is a conflict of interest between a decision-maker and a passive receiver of the consequences of the decision. The primary aim is to simulate the risk-taking externality that often (but not only) arises within financial and banking contexts with the aim of assessing whether a moral discussion between decision-makers and longer deliberation time can promote fairer choices for the passive receiver, thereby reducing the negative consequences associated with the risky choice. Indeed, within banking and finance, as much regulatory attention is now paid to ethical conduct as to prudential regulation. The application of ethical principles, such as fairness, to business behaviour is currently seen as a way to significantly affect both behaviour and mindset within financial organizations (i.e. culture), promoting a radical change in the financial sector (see for example De Nederlandsche Bank, 2015).

The results of our experiment support the hypothesis that moral discussion may reduce the risk-taking for others, while we do not find convincing evidence of the benefits of longer deliberation time. Thus, it seems that by only giving participants more time to think about their choice do not help them identify the relevant norm of behaviour. On the contrary, the frequency of unfair choices significantly decreases (with a relative risk between 0.50 and 0.70) in the treatments in which decision-makers can morally discuss the decision with peers (i.e. other participants facing the same decisions).

To better understand the driving mechanisms, we followed the standard practice of eliciting participants' first and second-order beliefs, i.e. expectations about behaviour of peers and expectations about other people's beliefs. Importantly, we observe that reduction in the frequency of unfair choices is accompanied by a significant change in both types of beliefs. In particular, we observe that in the treatments where a moral discussion takes place, decision-makers tend to believe that others will opt more often for the fair decision, and believe that others expect the same from them. No significant difference emerge after a reflection period. Additionally, we are able to disentangle between

[27] These results are robust at the different type of first-order beliefs that we use. If we use first-order belief for the same-room results are very similar. See Table (A.6) in Appendix A.

Table 9
Mediation Analysis: first order beliefs. Estimated Causal Effects of Moral Discussion.

	MDT vs RT	RT vs BT	MDT vs BT	Prosocial	Envy	Prosocial*Envy
Average Effect						
Direct - ADE	−0.018	−0.131	*−0.148*	−0.453	−0.356	−0.395
	[−0.151, 0.082]	[−0.279, 0.012]	*[−0.281 −0.039]*	*[−0.706 −0.156]*	*[−0.649 −0.069]*	*[−0.709 −0.092]*
Second Order Belief						
Mediation - ACME	0.021	−0.003	0.018	0.006	0.008	0.006
	[−0.022, 0.082]	[−0.035, 0.012]	[−0.017 0.073]	[−0.051 0.072]	[−0.040 0.079]	[−0.053 0.079]
First Order Belief						
Mediation - ACME	*−0.101*	−0.013	*−0.114*	−0.080	−0.086	−0.065
	[−0.181,−0.035]	[−0.096, 0.063]	*[−0.195 −0.550]*	[−0.294 0.095]	[−0.266 0.106]	[−0.256 0.119]
N obs	412					
Log-likelihood	−4314					

The outcome variable is *Unfair.* Each cell shows a point estimate and its corresponding 95% confidence intervals based on nonparametric bootstrap with 1000 resamples.

a direct and an indirect effect that a moral discussion has on decision makers' choices: a direct effect on the fair choices, and an indirect one that mainly unfolds through a modification of the empirical expectations (i.e first order beliefs). In other words, we observe that moral discussion increases the share of fair choices by helping participants better identify the consequences of their decisions.

Thus, although our set-up may appear relatively simplistic, and we cannot directly distinguish between the different types of norms at play in our experiment, e.g. social versus moral norms (Krupka and Weber, 2013; Schram and Charness, 2015; Bicchieri, 2016), the results from our experiment demonstrates that the moral discussion approach (the hallmark of ethics education and training) can be a powerful channel to convey a change within an organization. In particular, these results support the introduction of alternative – although not entirely "new" – methods to promote ethical behaviour among peers within an organization. In addition, even though this research has been inspired by the debate on the widespread dishonest behaviours in the financial industry (Cohn et al., 2014; Zingales, 2015), these results can be easily generalized to all situations in which there are risk externalities associated to individual choices.

To conclude, from the customers' side, the positive effect we observe on first-order beliefs suggests that consumers could trust more the investors' decisions once they know that investors will face moral discussions and opens an avenue for future research. Indeed, in future experiment, this intuition can be tested by using a trust game to investigate if a discussion between trustees improves trust of first movers.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jbef.2022.100735.

References

Diamond, D.W., Rajan, R.G., 2009. The credit crisis: Conjectures about causes and remedies. Amer. Econ. Rev. 99 (2), 606–610.

De Bruin, B., Herzog, L., O'Neill, M., Sandberg, J., 2018. Philosophy of money and finance.

Zingales, L., 2015. Presidential address: Does finance benefit society? J. Finance 70 (4), 1327–1363.

Mian, A., Sufi, A., 2017. Fraudulent income overstatement on mortgage applications during the credit expansion of 2002 to 2005. Rev. Financ. Stud. 30 (6), 1832–1864.

Eriksen, K.W., Kvaløy, O., 2017. No guts, no glory: An experiment on excessive risk-taking. Rev. Finance 21 (3), 1327–1351.

Linarelli, J., 2017. Luck, justice and systemic financial risk. J. Appl. Philos. 34 (3), 331–352.

James, A., 2017. The distinctive significance of systemic risk. Ratio Juris 30 (3), 239–258.

Morris, N., Vines, D., 2014. Capital Failure: Rebuilding Trust in Financial Services. OUP Oxford.

Guiso, L., Sapienza, P., Zingales, L., 2015. The value of corporate culture. J. Financ. Econ. 117 (1), 60–76.

Klooster, J., Meyer, M., 2016. Ethical banking: the key concepts. Mimeo,Cambridge Judge Business School.

Cohn, A., Fehr, E., Maréchal, M.A., 2014. Business culture and dishonesty in the banking industry. Nature 516 (7529), 86–89.

Cohn, A., Fehr, E., Maréchal, M.A., 2017. Do professional norms in the banking industry favor risk-taking? Rev. Financ. Stud. 30 (11), 3801–3823.

Egan, M., Matvos, G., Seru, A., 2019. The market for financial adviser misconduct. J. Polit. Econ. 127 (1), 233–295.

Suss, J., Bholat, D., Gillespie, A., Reader, T., 2021. Organisational culture and bank risk. Bank of England Working Paper.

Kanagaretnam, K., Lobo, G.J., Wang, C., Whalen, D.J., 2019. Cross-country evidence on the relationship between societal trust and risk-taking by banks. J. Financ. Quant. Anal. 54 (1), 275–301.

De Nederlandsche Bank, 2015. Supervision of behaviour and culture: Foundations, practice and future developments. De Nederlandsche Bank, Amsterdam, p. 14.

Fernández Muñiz, B., Montes Peón, J.M., Vázquez Ordás, C.J., 2018. Assessing and measuring banking culture. In: García-Olalla, M., Clifton, J. (Eds.), Contemporary Issues in Banking: Regulation, Governance and Performance. Springer International Publishing, Cham, pp. 363–387.

Wehinger, G., et al., 2012. Banking in a challenging environment: business models, ethics and approaches towards risks. OECD J. Financial Market Trends 2, 79–88.

D'Acunto, F., 2018. Tear down this wall street: Anti-finance rhetoric, subjective beliefs, and investment. In: Subjective Beliefs, and Investment. July 14, 2018.

Reynolds, J.N., Newell, E., 2011. Ethics in Investment Banking. Springer.

Lo, A.W., 2015. The Gordon Gekko effect: The role of culture in the financial industry. Technical Report, National Bureau of Economic Research.

Young, H.P., Noe, T., et al., 2012. The Limits to Compensation in the Financial Sector. Technical Report, Department of Economics Oxford.

Awrey, D., Blair, W., Kershaw, D., 2013. Between law and markets: Is there a role for culture and ethics in financial regulation? Del. J. Corp. Law (DJCL) 38, 191.

Kirchler, M., Lindner, F., Weitzel, U., 2018. Rankings and risk-taking in the finance industry. J. Finance 73 (5), 2271–2302.

Moore, D.A., Loewenstein, G., 2004. Self-interest, automaticity, and the psychology of conflict of interest. Soc. Justice Res. 17 (2), 189–202.

Bénabou, R.J., Falk, A., Tirole, J., 2018. Eliciting Moral Preferences. Mimeo.

Gert, B., Gert, J., 2017. The definition of morality. In: Zalta, E.N. (Ed.), The Stanford Encyclopedia of Philosophy, Fall 2017 Metaphysics Research Lab, Stanford University.

Chakravarty, S., Harrison, G.W., Haruvy, E.E., Rutström, E.E., 2011. Are you risk averse over other people's money? South. Econ. J. 77 (4), 901–913.

Dana, J., Weber, R.A., Kuang, J.X., 2007. Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness. Econom. Theory 33 (1), 67–80.

Cason, T.N., Mui, V.-L., 1997. A laboratory study of group polarisation in the team dictator game. Econ. J. 107 (444), 1465–1483.

Bolton, G.E., Ockenfels, A., Stauf, J., 2015. Social responsibility promotes conservative risk behavior. Eur. Econ. Rev. 74, 109–127.

Eijkelenboom, G.G., Rohde, I., Vostroknutov, A., 2019. The impact of the level of responsibility on choices under risk: the role of blame. Exp. Econ. 22 (4), 794–814.

Haidt, J., 2001. The emotional dog and its rational tail: a social intuitionist approach to moral judgment. Psychol. Rev. 108 (4), 814.

Imai, K., Keele, L., Tingley, D., 2010a. A general approach to causal mediation analysis. Psychol. Methods 15 (4), 309.

Imai, K., Keele, L., Yamamoto, T., 2010c. Identification, inference and sensitivity analysis for causal mediation effects. Statist. Sci. 51–71.

Hillenbrand, A., Verrina, E., 2019. The differential effect of narratives on prosocial behavior. MPI Collective Goods Discussion Paper, (2018/16).

Fehr, E., Schurtenberger, I., 2018. Normative foundations of human cooperation. Nat. Hum. Behav. 2 (7), 458.

Bicchieri, C., 2006. The Grammar of Society: The Nature and Dynamics of Social Norms. Cambridge University Press.

Krupka, E.L., Weber, R.A., 2013. Identifying social norms using coordination games: Why does dictator game sharing vary? J. Eur. Econom. Assoc. 11 (3), 495–524.

Kallgren, C.A., Reno, R.R., Cialdini, R.B., 2000. A focus theory of normative conduct: When norms do and do not affect behavior. Pers. Soc. Psychol. Bull. 26 (8), 1002–1012.

Cialdini, R.B., Reno, R.R., Kallgren, C.A., 1990. A focus theory of normative conduct: recycling the concept of norms to reduce littering in public places. J. Personal. Soc. Psychol. 58 (6), 1015.

Gunia, B.C., Wang, L., Huang, L., Wang, J., Murnighan, J.K., 2012. Contemplation and conversation: Subtle influences on moral decision making. Acad. Manag. J. 55 (1), 13–33.

Gneezy, U., 2005. Deception: The role of consequences. Am. Econ. Rev. 95 (1), 384–394.

Andersen, S., Gneezy, U., Kajackaite, A., Marx, J., 2018. Allowing for reflection time does not change behavior in dictator and cheating games. J. Econ. Behav. Organ. 145, 24–33.

Greene, J.D., 2014. Beyond point-and-shoot morality: Why cognitive (neuro) science matters for ethics. Ethics 124 (4), 695–726.

Moore, C., Tenbrunsel, A.E., 2014. "Just think about it?" cognitive complexity and moral choice. Organ. Behav. Hum. Decis. Process. 123 (2), 138–149.

Kohlberg, L., 1976. Moral stages and moralization. Moral Dev. Behav. 31–53.

Krajbich, I., Bartling, B., Hare, T., Fehr, E., 2015. Rethinking fast and slow based on a critique of reaction-time reverse inference. Nature Commun. 6, 7455.

Frank, D.-A., Chrysochou, P., Mitkidis, P., Ariely, D., 2019. Human decision-making biases in the moral dilemmas of autonomous vehicles. Sci. Rep. 9 (1), 1–19.

Bicchieri, C., Dimant, E., 2019. Nudging with care: The risks and benefits of social information. Public Choice 1–22.

Habermas, J., 1990. Moral Consciousness and Communicative Action. MIT Press.

Jia, J., Dyer, J.S., 1996. A standard measure of risk and risk-value models. Manage. Sci. 42 (12), 1691–1705.

Bartling, B., Fehr, E., Maréchal, M.A., Schunk, D., 2009. Egalitarianism and competitiveness. Am. Econ. Rev. 99 (2), 93–98.

John, O.P., Donahue, E.M., Kentle, R.L., 1991. The big five inventory: versions 4a and 54. University of California, Berkeley, Institute of Personality and Social Research, Berkeley, CA.

John, O.P., Naumann, L.P., Soto, C.J., 2008. Paradigm shift to the integrative big five trait taxonomy. Handbook Pers. Theory Res. 3, 114–158.

Kocher, M.G., Sutter, M., 2005. The decision maker matters: Individual versus group behaviour in experimental beauty-contest games. Econ. J. 115 (500), 200–223.

Sutter, M., 2009. Deception through telling the truth?! experimental evidence from individuals and teams. Econ. J. 119 (534), 47–60.

Hoenig, J.M., Heisey, D.M., 2001. The abuse of power: the pervasive fallacy of power calculations for data analysis. Amer. Statist. 55 (1), 19–24.

Lakens, D., 2017. Equivalence tests: A practical primer for t tests, correlations, and meta-analyses. Soc. Psychol. Pers. Sci. 8 (4), 355–362.

Schram, A., Charness, G., 2015. Inducing social norms in laboratory allocation choices. Manage. Sci. 61 (7), 1531–1546.

Battigalli, P., Dufwenberg, M., 2007. Guilt in games. Amer. Econ. Rev. 97 (2), 170–176.

Bicchieri, C., Xiao, E., 2009. Do the right thing: but only if others do so. J. Behav. Decis. Mak. 22 (2), 191–208.

Kimbrough, E.O., Vostroknutov, A., 2016. Norms make preferences social. J. Eur. Econom. Assoc. 14 (3), 608–638.

Bicchieri, C., 2016. Norms in the Wild: How to Diagnose, Measure, and Change Social Norms. Oxford University Press.