Variations in the Two-Step Norm Elicitation Procedure

Sabrina Plaß (Corresponding author)
Management Department, Paderborn University,
Fürstenallee 11
33102 Paderborn, Germany
sabrina.plass@uni-paderborn.de
ORCID: 0000-0001-9867-8079

The two-step norm elicitation procedure describes a commonly used tool for measuring normative expectations in an incentivized way. This study tests some of its design features to determine whether elicited beliefs and related behaviours vary depending on i) the time of elicitation (before vs after the decision), ii) incentivizing vs not incentivizing a question about normative expectations, and iii) questioning subjects on their beliefs about the action of interest alone or combined with an alternative action. An online experiment is conducted via Prolific comprising a dictator game and the elicitation of fairness beliefs. A pretest reveals that applying role uncertainty does not alter beliefs and behaviours compared to a baseline treatment without it. Subsequently, three treatments are implemented. Contrary to previous studies, results indicate that varying the time of elicitation does not significantly alter the money-share decision. However, incentivizing the question about normative expectations significantly increases the fit with the actual majority norm. Finally, asking about a fair share and an unfair share instead of only about fair sharing does not alter personal normative beliefs or normative expectations, but it increases the empirical expectations that other dictators have provided a fair share.

Keywords: *Social Norms, Normative Expectations, Personal Normative Belief, Elicitation, Economic Experiment*

Declarations

The authors have no relevant financial or non-financial interests to disclose.

1. Introduction & Background

The influence of *social norms* on behaviour has sparked increasing interest in economics in the last decades. Social norms are generally described as rules about what is and is not appropriate behaviour (Görges & Nosenzo, 2020). In contrast to other-regarding preferences, we talk of a social norm when a behaviour is influenced by the dominant beliefs of a person's reference group – which can be assumed (normative expectation) or known (empirical expectation) (Bicchieri, 2006). Social norms are enforced either externally, through social punishments (e.g., avoidance, ostracism), or material sanctions (e.g., loss of property), or through internalised sanctions (e.g., feelings of guilt) (e.g., Bicchieri et al., 2018; Cialdini & Goldstein, 2004; Fehr & Fischbacher, 2004). Previous studies on social norms enhance economic theories by demonstrating that individuals do not merely act out of self-interest but also consider whether others support a particular behaviour, or behave in a particular situation (Bicchieri, 2006; Fehr & Schmidt, 2006). Hence, considerations of the utility of social approval and the potential impact on social belonging complement purely monetary considerations. The study of social norms often involves standard economic experiments, such as the dictator game (e.g., Bicchieri & Xiao, 2009; Gächter et al., 2017; Kimbrough & Vostroknutov, 2018). Experimental evidence suggests that social norms impact decision-making, especially for pro-social behaviour enacting a sense of fairness (Gächter et al., 2017; Krupka & Weber, 2013), charitable giving (Agerström et al., 2016), and in the context of honesty/lying (Abeler et al., 2019; Bicchieri et al., 2023), corruption (Gneezy et al., 2019), cooperative behaviour (Reuben & Riedl, 2013), whistleblowing (Mir Djawadi et al., 2025), and discrimination against out-group members (Barr et al., 2018).

According to Bicchieri (2006, 2017), there are three types of beliefs. Personal normative beliefs describe an individual's belief about the right course of action in a given situation (first-order beliefs). Social expectations concern one's beliefs about expected behavioural norms and the corresponding behaviour of others (Bicchieri, 2017). More specifically, normative expectations describe one's perception of the majority belief of a reference group about what one *ought to do* in a certain situation, and is composed of the sum of the personal normative beliefs of others, and referred to as second-order beliefs (Bicchieri 2017). By contrast, empirical expectations are non-normative and refer to expectations about how one perceives the majority of a reference group to behave in a given situation. Hence, social norms represent second-order beliefs, and behavioural choices are guided by both normative and empirical expectations (Bicchieri, 2006).

The increasing importance of social norms in economics lends emphasis to the need for robust measurements to advance empirical research, and specifically to test theories of how social norms translate into behaviour. The most common approaches for measuring social norms include the non-incentivized 'belief survey' method, the incentivized 'Krupka-Weber' method (2013) for eliciting second- (or higher-)order beliefs, and the two-step elicitation method by Bicchieri and Xiao (2009),

also called 'opinion matching' method, eliciting personal normative beliefs and subsequently normative expectations (Görges & Nosenzo, 2020; Lane et al., 2023). Görges and Nosenzo (2020) summarize and critically reflect upon these approaches and point out that, while neither method is superior per se, each has advantages and drawbacks, and may suit different research questions. Although the Krupka-Weber method has gained much traction lately, the two-step elicitation method by Bicchieri and Xiao (2009) continues to provide some distinct advantages over the Krupka-Weber method. First, it eliminates the latter's strategy component (no distortion of beliefs due to strategic coordination), second, it elicits both personal normative beliefs and normative expectations.[1] This is especially useful because the two-step elicitation method could reveal potential discrepancies between, on the one hand, personal normative beliefs and, on the other, normative expectations, and in so doing it may also uncover the mechanisms driving the misperception (under- or overestimating the majority's norm), or even pluralistic ignorance – the false assumption that one's own personal beliefs differ from those of the majority (Bicchieri et al., 2023; Bursztyn, González, et al., 2020; Sargent & Newman, 2021).

A number of authors have applied the two-step elicitation method (e.g., Bicchieri et al., 2020, 2021, 2023; Bogliacino et al., 2024; Bursztyn, Egorov, et al., 2020; Bursztyn, González, et al., 2020; D'Adda et al., 2020; Lane et al., 2023; Mir Djawadi et al., 2025). However, while various robustness checks have been applied to the Krupka-Weber method (e.g., Castillo et al., 2022; D'Adda et al., 2016; Fallucchi & Nosenzo, 2022; König-Kersting, 2024), only a few have been carried out on the two-step elicitation method (e.g., Aycinena et al., 2024 testing potential social desirability biases). For scholars intending to adopt the two-step elicitation method, it might become essential to pay close attention to its specific design features, for example, whether eliciting norms before or after the decision might impact decision differently.

The aim of this study is to test the robustness of variations in the two-step elicitation method, and whether (and how) these variations affect normative expectations and behaviour. We vary three different components: i) time of elicitation (before vs after the decision), ii) incentivization of normative expectations (incentivized vs non-incentivized)[2] and iii) eliciting beliefs about one vs at least two behavioural alternatives. These variations are tested in an online experiment deploying a variation of the dictator game (using role uncertainty) with UK participants recruited from the platform Prolific, and

[1] Krupka and Weber (2013)'s method elicits only normative expectations (using a Likert scale). Participants are paid a monetary reward if their rating matches the majority rating by others. However, as they are incentivized to choose the rating, they believe most others will choose, it's not even clear whether the Krupka-Weber method elicits second-order beliefs (normative expectations) or higher-order beliefs (participants' expectations about what others believe is the normative expectation) (Görges & Nosenzo, 2020).

[2] A critical aspect of the method is that the non-incentivization of personal normative beliefs could translate into a distortion of first-order belief, e.g., through a response bias. This, however, has been partly cancelled out by Aycinena et al. (2024), who ran several experiments and concluded that the elicitation method is not in itself prone to a social desirability bias.

follows the general gist of Bicchieri and Xiao (2009) to elicit norms about fair sharing[3]. We test the robustness of the elicitation method in this context because previous findings have already demonstrated that fairness is a social norm (Bicchieri & Xiao, 2009). A plethora of studies using a dictator game indicates that individuals do not merely act selfishly when dividing budgets but weigh up their own payoff maximization against their social obligations (Engel, 2011). For instance, scholars indicate that dictators seem to consider what they believe is a fair share guided by personal and social rules (Bolton et al., 1998), and by the rule of reciprocity (Camerer & Thaler, 1995), while Krupka and Weber (2013) revealed that participants base their utility on taking actions that they perceive as socially appropriate in terms of a fair share.

This study contributes with four main insights to the literature on social norms and dictator games. First, we find across all treatments, that individuals believe others find a fair share less appropriate as they do. Second, behaviour does not significantly differ when eliciting normative expectations before the task compared to the elicitation after the task. Third, incentivization significantly increases the accuracy of normative expectations (that means the correct estimation of the majority norm). Finally, asking about two behavioural alternatives – fair share and low share – increases empirical expectations of the (social norm) behaviour.

2. Predictions & Experimental Design

2.1. Predictions

2.1.1. Prediction concerning the Time of Elicitation

Based on the literature reviewed above, several predictions will be made concerning three main variations in the two-step norm elicitation procedure: i) the time of elicitation, ii) the incentivization of normative expectations, and iii) elicited beliefs on behavioural alternatives. The first of these is the time of belief elicitation, which could influence how participants behave in a task. Eliciting beliefs before the task is especially useful to investigate the impact of social information interventions on behaviour[4]. However, according to Bicchieri (2006), the salience of a norm may increase compliance with it (Bicchieri, 2006; Horne & Mollborn, 2020). Several studies have found that eliciting beliefs before the task could impact the behaviour. Evidence for this was found in experiments, for example, Mir Djawadi et al. 2015 (whistleblowing experiment), in Gächter and Renner 2010 (public goods experiment), Bicchieri and Chavez 2010 (ultimatum game). The assumption is that norm elicitation before the task

[3] We follow the terms used in Bicchieri and Xiao and refer to a fair sharing for the dictator dividing the budget equally. Note that in dictator games this is often described as generosity (Engel, 2011).

[4] In experiments that investigate the influence of social information interventions, participants are usually first asked about their beliefs, and then given the information, that is, normative or descriptive messages about, for example, the majority belief from previous treatments before making a decision.

could induce a framing effect and trigger the activation of social norms by making the normativity of the decision more salient. Brañas-Garza (2007) shows that behaviour in the dictator game can be prone to framing and demand effects[5], while Dreber et al. (2013) did not find support for a (far less demanding) framing effect. We propose that merely asking about personal normative beliefs and normative expectations might have a similar effect as in Brañas-Garza (2007) and nudge behaviour. Therefore, we predict as follows:

Prediction P1: Eliciting beliefs and expectations before the decision increases fair sharing compared to elicitation after the decision.

2.1.2. Prediction concerning the Incentivizing of the Elicitation

Studies and empirical evidence on whether or not to incentivize the elicitation of beliefs are somewhat controversial. In an ultimatum game using the Krupka-Weber method, Veselý (2015) did not find a significant difference between incentivizing and non-incentivizing the elicitation of beliefs. By contrast, incentivizing the perception of normative expectations significantly increased accuracy in a public goods game (Gächter & Renner, 2010). Moreover, it is reasoned that incentivization reduces automatic thinking (System 1 thinking) and encourages more effortful thinking (Epley & Gilovich, 2005). Accordingly, we predict that incentivizing the elicitation of normative expectations leads to participants thinking more carefully about what the most common response could be, because the incentivization increases the salience of the normative expectations. Hence, we formulate our second prediction:

Prediction P2: Incentivizing the elicitation of normative expectations increases the accuracy (correct estimate of the majority norm) compared to not incentivizing the elicitation.

2.1.3. Predictions concerning Questions about a Fair Share and an Alternative (Low) Share

Lastly, we are interested in whether there is a difference if participants are asked about their personal normative belief and their normative expectations of the behaviour of interest only compared to being asked about at least two behaviours. In other words, we are interested in whether questions about alternative behaviours distort the perception of beliefs. As alternative share we refer to a low share, which will be investigated as the second option. Asking about the beliefs of at least two (mutually exclusive) behavioural decisions has the potential to identify whether multiple normative expectations influence behaviour. Do individuals have consistent and unambiguous beliefs and expectations in the sense that they find behaviour A being appropriate and behaviour B inappropriate? Or do individuals

[5] In his study a sentence "Note that the receiver relies on you" influenced behaviour, which however, can be regarded as an induced social rule and a demand effect by the experimenter.

hold multiple inconsistent or ambiguous normative expectations by expecting two (mutually exclusive) behaviours to be similarly appropriate (or inappropriate)? Or is the behaviour even subject to polarized norms (opposing views that reinforce group divisions)?

For the context of this study, empirical evidence supports that fairness is a known and unambiguous social norm (e.g., Bicchieri & Xiao, 2009; Engel, 2011). Therefore, assessing beliefs about other behaviours should not distort the fairness norms, only provide additional information. Hence, we do not expect any difference in personal normative belief and normative expectations of the appropriateness of fair sharing when asking about more behavioural options. In a similar vein, we do not expect that empirical expectations, the belief that other provide a fair share in our context, alter significantly either. Therefore, we assume that asking questions about the appropriateness of at least two behaviours instead of only one might affect the beliefs as follows:

Prediction P3.1: There is no difference in the appropriateness concerning personal normative beliefs of fair sharing when asking for beliefs for both behaviours, compared to eliciting personal normative beliefs about one behaviour alone.

Prediction P3.2: There is no difference in the appropriateness concerning normative expectations of fair sharing when eliciting expectations for both behaviours, compared to eliciting normative expectations about one behaviour alone.

Prediction P3.3: The number of participants having empirical expectations that others provide a fair share does not alter when eliciting normative expectations for both behaviours, compared to eliciting normative expectations concerning one behaviour alone.

2.2. Basic Experimental Set-up

The experiment mostly follows the structure of Bicchieri and Xiao (2009)'s applied dictator game. As the dictator game can be very sensitive to design elements (e.g., Brañas-Garza, 2007), we closely follow Bicchieri and Xiao's design in terms of the instructions and the increments for the money shares. However, some changes have been made, as will be explained. Instead of an on-site pen-and-paper game, UK participants were recruited via the Prolific platform to play an online version of the game. Dictators, called dividers in the instructions, have the task of dividing £2.50 between themselves and a receiver. For their share they can only choose one of the options A-G (see Figure 1). Options C and D, which give the receiver £1 or £1.25, respectively (between 40% and 50% of the total budget), are referred to as 'fair shares'. Options A and B, which give the receiver 25p (£0.25) or 50p (between 20% and 30% of the total budget), are referred to as 'low shares'[6], and options E-G (from £1.50 to £2.25) as 'high shares'. Dictators can choose to allocate the budget in 25p increments, excluding the split of £1.75

[6] Bicchieri and Xiao (2009) refer to these options as selfish shares; however, this was changed to low shares, to stay neutral in language.

to the dictator and 75p to the receiver (exactly between a fair and a low share) and the split of 75p to the dictator and £1.75 to the receiver (exactly between a fair and a high share).

Figure 1: Dividing Options

Possible options	The split
A	Dictator gets **£2.25** and Receiver gets **£0.25**
B	Dictator gets **£2.00** and Receiver gets **£0.50**
C	Dictator gets **£1.50** and Receiver gets **£1.00**
D	Dictator gets **£1.25** and Receiver gets **£1.25**
E	Dictator gets **£1.00** and Receiver gets **£1.50**
F	Dictator gets **£0.50** and Receiver gets **£2.00**
G	Dictator gets **£0.25** and Receiver gets **£2.25**

This study's design also differs by applying role uncertainty. Each participant starts off in the role of dictator and has to decide on an option[7]. After having made their decision, participants are randomly paired, with one selected to be the dictator, and the other the receiver. The dictator's chosen allocation will then be implemented.

Before proceeding with the experiment description, an explanation is due on our design's use of role uncertainty. As mentioned, scholars have previously indicated that dictator games are sensitive to design modifications, and several variations have been tested since (Camerer, 2003; Cox, 2010; Fehr & Schmidt, 2006; Kassas & Palma, 2019; Walkowitz, 2021). For instance, Heinrich & Weimann (2013) show that the dictator's behaviour is not influenced when the payoff-relevant game is chosen by the recipients compared to a random assignment. However, role uncertainty has been indicated as a potential drawback because it influences behaviour (e.g. Iriberri & Rey-Biel, 2011; Mesa-Vázquez et al., 2021; Walkowitz, 2021). Therefore, a pretest is conducted to test whether role uncertainty does alter the results from the baseline treatment. Results, more precisely depicted in Section 3.1, indicate no difference when applying role uncertainty.

[7] This is not to be confused with the strategy method (Selten, 1967), where, for a task with a first and a second mover, all possible options can be observed. By contrast, this study does not involve a second mover, but in order to gain as many observations as possible, dictator decisions are collected from all participants. Participants receive information on whether they or their teammate is in the role of the dictator after the study, which is commonly referred to as role uncertainty.

Table 1: Description of all the Stages in the Baseline Experiment T(Base), excluding Treatment Variations

Timeline of a session in T(Base).	
Sequence of activities	**Activity description**
1	Participants read a short description of the study on the Prolific platform and decide to participate.
2	Participants enter the experiment; **session starts.**
3	Participants read **general instructions.** Instructions provide complete information about of the experiment.
4	**Experiment** starts.
5	**Dictator** decision (all participants).
6	Elicitation of **personal** normative beliefs and **normative expectations** (incentivized).
7	Elicitation of **empirical expectations**.
8	**Experiment ends; Survey starts.**
9	**Session ends**.
10	Participants are **paid** the fixed payment (£1) anonymously via the Prolific platform.
11	Participants are randomly **matched**; randomly one is chosen to be dictator and the other receiver; the decision of the dictator will be implemented.
12	Participants are paid an additional **bonus payment** according to the decision of the dictator and if they are correct in the elicitation of the appropriateness.

After subjects have decided on the budget share, the experiment elicits participants' personal normative beliefs and normative expectations (the most common response) regarding the fair share (Option C or D).[8] Participants are asked to select whether they personally believe providing a fair share is very inappropriate, somewhat inappropriate, somewhat appropriate or very appropriate. To elicit their normative expectations, participants are asked which answer they think most of the other participants did choose in the preceding question, presenting them with the same four response options (very inappropriate – very appropriate). Normative expectations are incentivized in that participants will

[8] This is a deviation from Bicchieri and Xiao (2009), who ask personal normative beliefs as a dummy variable ("Do you think you should make a fair offer? / Do you think that dividers should split the money approximately equally (choose C or D)?) and elicit a concrete number for normative expectations ("How many dividers in this room do you think answered 'Yes' to question (d)?"). Instead, to account for more nuances, this study employs a Likert scale for participants to indicate the appropriateness of the actions. This adaption was needed because giving a concrete number in the online experiment was not feasible due to potential dropouts. The formulation of the questions is similar to that used by Krupka and Weber (2013), but the specification "socially" in front of 'appropriate' has been dropped.

receive an additional bonus of 25p if their answer to the question matches the correct majority choice. On the next screen, participants are asked about empirical expectations, more specifically, whether they expect others to provide a fair share (yes/no). Bicchieri (2006) emphasizes the importance of the reference group in the context of social norms. Hence, our participants are informed that all the other participants in this study are UK residents. Additionally, it is explained that the term "inappropriate" means an unacceptable, unsuitable or incorrect action in that situation. The term "appropriate" indicates an acceptable, suitable or correct action in that situation.

Lastly, participants are asked to complete a questionnaire about their demographics, control variables, fairness items of the Moral Foundation Questionnaire (Graham et al., 2011), and two attention check questions. In a baseline treatment T(Base), the experiment is conducted as explained above and depicted in Table 1. A pretest is conducted to check whether role uncertainty distorts beliefs and behaviours compared to T(Base).

2.3. Treatment Variations

Three treatments will be conducted implementing different variations of the elicitation procedure (see Figure 2). In treatment T(Before), the elicitation time varies. Everything else is held constant, but the elicitation of beliefs and expectations is conducted before the decision, while in T(Base), the elicitation takes place after the decision. In treatment T(NoIncentive), the additional incentivization for correctly stating the majority norms is dropped, whereas in T(Base), participants are externally incentivized to carefully consider the question to receive a bonus payment (if their estimation equals the most common response). All other aspects are held constant. In treatment T(Questions), in addition to being asked about the appropriateness of a fair share, participants are also asked about their personal normative beliefs and the normative expectation (incentivized) of a low share (Option A or B)[9].

Figure 2: Treatment Variations / Manipulations

	T(Base)	**T(Before)**	**T(NoIncentives)**	**T(Questions)**
Time of elicitation	after	before	after	after
Incentivized elicitation	yes	yes	no	yes
Questions about a low and a fair share	fair share	fair share	fair share	both

[9] The question about personal normative beliefs of a low share states: "How appropriate do you personally believe it is to make a low offer to the Receiver (Option A or B)?", with the potential answers ranging from very inappropriate, somewhat inappropriate, somewhat appropriate to very appropriate. The question on normative expectations asks: "Which answer do you think the majority of participants chose in the preceding question? If your answer matches the actual answer of the majority, you will earn an additional 25p."

An a-priori sample size calculation, assuming a small to medium-sized effect (Cohen's d: 0.35; probability level: 0.05; statistical power level: 0.8), reveals that 102 observations per treatment are needed. Each participant is assigned to only one of the treatments.

2.4. Procedure

The experiment was programmed using oTree (Chen et al., 2016) and conducted in December 2024 via the platform Prolific Academic[10] with participants from the UK. There are three main reasons why UK residents were chosen as participants, one of which was the requirement to be fluent in English (criteria set in Prolific). Second, recruiting participants from the same country provides them with a reference group they can more easily refer to as they share at least broadly the same institutional and cultural background. Third, Prolific originates from the UK and has the most participants there. The study received prior ethical approval, and the study has been pre-registered[11]. Participants received the instructions at the beginning of the experiment[12].

3. Results

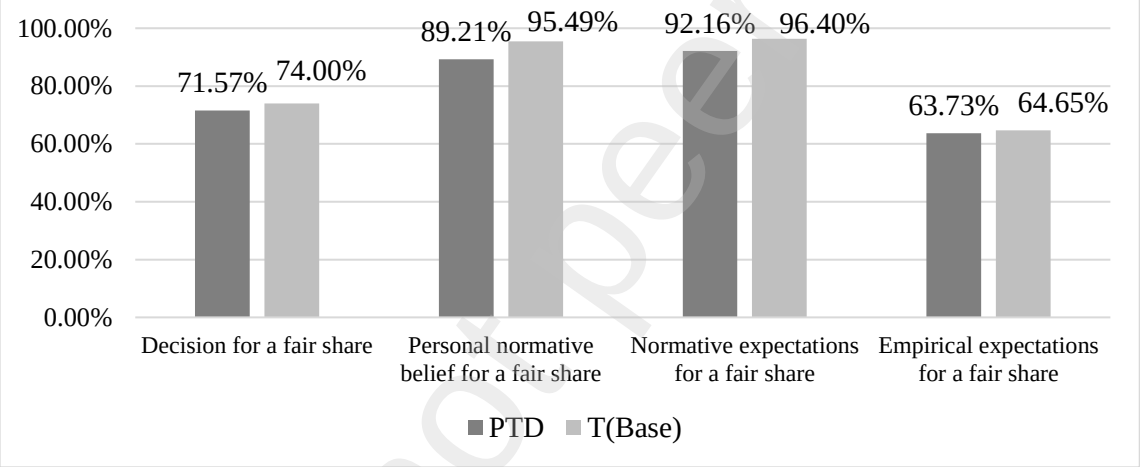
3.1.Pretest

Before conducting the treatments, we run a pretest to investigate whether role uncertainty influences behaviour and beliefs in the studies' experimental design. The pretest and its findings are briefly described before reporting the main results. The pretest aims to investigate whether the results differ from those using the standard procedure in T(Base). In T(Base) all participants perform the task as dictators and, to implement role uncertainty, they are randomly assigned to either the role of *dictator* (called divider in experiment*)* or *receiver* only at the end. By contrast, the pretest performs a dictator game without using role uncertainty, where participants are informed of their role of dictators or receivers in the study beforehand (divided for analysis into PTD=Pretest Dictator and PTR=Pretest Receiver). The share provided by the dictators in PTD, and the norms elicited from them, are then compared to T(Base).

[10] We deliberately chose Prolific in comparison to other crowdworking platforms. Participants on Prolific have been found to be more naïve and less dishonest than, say MTurk (Peer et al., 2017); Prolific allows the inclusion of specific requirements for participants (Peer et al., 2017), and Prolific participants show lower levels of attentional disengagement than MTurk participants (Albert & Smilek, 2023). For further advantages and functionalities compared to other platforms refer to Palan and Schitter (2018).
[11] For the ethical approval from the GfeW see: https://gfew.de/ethik/Bm3XqJ6f and for the peer-reviewed re-registration of the study see Social Science Registry: https://www.socialscienceregistry.org/trials/14794.
[12] For the instructions see online appendix: https://t1p.de/ci7h0.

In total, 201 subjects participated in the pretest with 102 as dictator (PTD) and 99 as receiver (PTR, some failed attention checks or did not finish), and 111 in T(Base). The average amount given to receivers is £0.95 in PTD and £1.00 in T(Base). Additionally, the mean value of elicited beliefs is similar, with personal normative beliefs on average measuring 3.42 in PTD and 3.55 in T(Base), and normative expectations 3.27 in PTD and 3.35 in T(Base). Figure 3 depicts the percentage differences between dictators of PTD and T(Base) (for a more detailed analysis, see Table A1 in Appendix). Results reveal no significant difference in personal normative beliefs about a fair share (Chi-Square Test: $\chi 2$ (3) = 3.1156, p = 0.374), no significant difference between normative expectations ($\chi 2$ (3) = 1.9994, p = 0.573) and no difference in empirical expectations ($\chi 2$ (1) = 0.0100, p = 0.920). Moreover, the decision for a fair share did not differ significantly (two-sided Mann-Whitney U Test: z = -0.561, p = 0.5745). Therefore, the subsequent treatments use role uncertainty in the dictator game, where each participant is one observation unit.

Figure 3: Descriptive Data of the Pretest (Percentage Differences between PTD and T(Base)



3.2. Descriptive Data

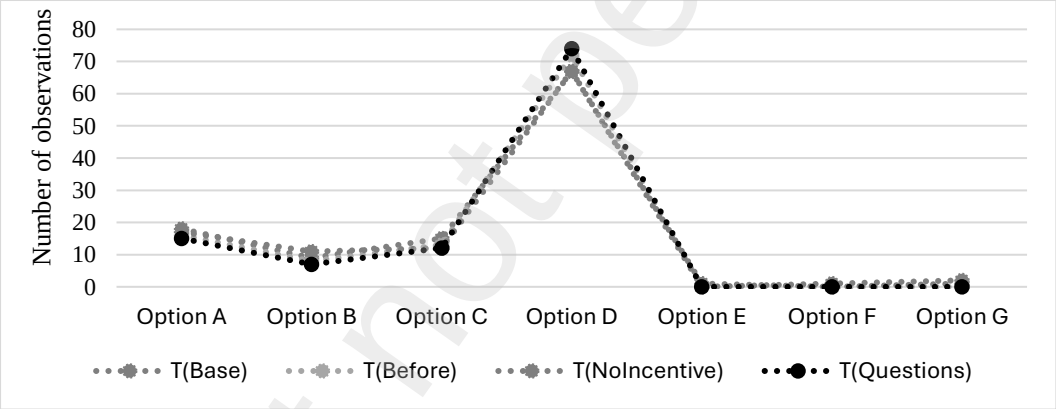
Across all treatments, 28 participants who did not finish the study or failed one of the attention checks were excluded. The final sample totals 438 subjects, of which 111 are in (T(Base), 109 in T(Before), 110 in T(NoIncentive) and 108 in T(Questions). More than 99% of participants found the instruction comprehensible. Participants received a fixed payment of £1 and on average an additional £1.38 as a bonus payment (on average in T(Base): £1.35, in T(Before): £1.37, in T(NoIncentive): £1.33, and in T(Questions): £1.48)[13].

[13] This aligns with the average bonus payoff in the Pretest, which amounts to £1.36.

There is no significant difference in the treatment compositions concerning age (Kruskal-Wallis Test: χ^2 (3) = 2.255, p = 0.5212), gender (Chi-Square Test: χ^2(6) = 7.5728, p = 0.271), education (χ^2 (21) = 13.0407, p = 0.907), whether or not participants are in employment (χ^2 (6) = 4.4317, p = 0.618), or which role they have at work (χ^2 (39) = 34.8594, p = 0.659). Moreover, there is no significant difference between the share in the decision task (χ^2 (18) = 11.8959, p = 0.853). The majority in each treatment decided on Option D, that is, one defined as a fair share (see Table 2, Figure 5 and Table A1 in the Appendix).

The majority of dictators across all treatments share the budget equally (Figure 4). More precisely, dictators granted receivers on average about 40% of the money (mean in T(Base): £1.00, in T(Before): £1.02, in T(NoIncentive): £1.02, and in T(Questions): £1.03).[14] These results are in line with previous studies inferring participants are not profit-maximizing but hold norms for fairness and generosity. In a meta-analysis, Engel (2011) finds that across numerous dictator games (616 treatments from 129 studies) dictators provide on average 28%. Several other studies show that the majority provides an equal share in the dictator game and expects this to be normatively the right thing to do (e.g., Andreoni & Bernheim, 2009; Kimbrough & Vostroknutov, 2016, 2018; Krupka & Weber, 2013).

Figure 4: Dictators Behaviour across Treatments



Note: Option A and Option B were classified as a low share, Options C and D as fair share, and Options E to G as high shares.

On average, 90% of the participants hold the personal normative belief that a fair share is (somewhat or very) appropriate (Table 2, Figure 5). This aligns with normative expectations, where on average 90% expect a fair share to be appropriate. Interestingly, even though both normative expectations and personal normative beliefs are regarded as (somewhat or very) appropriate by the majority, the degree of appropriateness is distributed in the opposite way: the majority expect others to find a fair share only *somewhat* appropriate (52.22%) but have personal normative beliefs of a fair share

[14] We cannot directly compare these results with Bicchieri and Xiao (2009) because they do not have a neutral or baseline treatment, but all treatments contain a specific norm message and are compared with one another.

being *very* appropriate (59.26%). By contrast, 38.88% expect the normative expectation of a fair share to be *very* appropriate, while 33.89% personally believe a fair share is *somewhat* appropriate.

Table 2: Descriptive Data of Dictators

Behaviour	**Low share**	**Fair share**	**High share**	**Share = majority norm** (D in all treatments)
T(Base)	24.33%	73.87%	1.8%	60.36%
T(Before)	24.02%	77.98%	/	60.91%
T(NoIncentive)	25.45%	71.82%	2.73%	68.51%
T(Questions)	20.37%	79.63%	/	62.41%

Empirical expectations	**Fair share (yes)**	**Fair share (no)**
T(Base)	63.06%	36.94%
T(Before)	69.72%	30.28%
T(NoIncentive)	62.73%	37.27%
T(Questions)	74.07%	25.93%

Personal normative beliefs about a fair share	**Very inappropriate**	**Somewhat inappropriate**	**Somewhat appropriate**	**Very appropriate**
T(Base)	2.70%	1.80%	33.33%	62.16%
T(Before)	0.92%	1.83%	36.70%	60.55%
T(NoIncentive)	1.82%	10%	28.18%	60%
T(Questions)	2.78%	1.85%	41.67%	53.70%

Normative expectations about a fair share	**Very inappropriate**	**Somewhat inappropriate**	**Somewhat appropriate**	**Very appropriate**
T(Base)	1.80%	1.80%	55.86%	40.54%
T(Before)	1.83%	4.59%	43.12%	50.46%
T(NoIncentive)	1.82%	15.45%	54.55%	28.18%
T(Questions)	1.85%	7.41%	54.55%	35.19%

The according mean values reinforce that normative expectations are lower than personal normative beliefs across the treatments. T(Base) revealed a mean value of personal normative beliefs of 3.55 (sd = 0.06) for the appropriateness of a fair share, and values in other treatments are similar (3.57 (sd = 0.06) in T(Before), 3.46 (sd = 0.07) in T(NoIncentive), and 3.46 (sd = 0.07) in T(Questions)). Additionally, the mean value for normative expectations of a fair share is similarly distributed across treatments with a value of 3.35 (sd = 0.06) in T(Base), 3.42 (sd = 0.07) in T(Before), 3.09 (sd = 0.07) in T(NoIncentive), and in T(Questions) it is 3.24 (sd = 0.08). Additional analysis supports the finding that personal normative beliefs significantly differ from normative expectations in all treatments. As mentioned, while the majority of participants find a fair share very appropriate, the majority expects that others hold the belief that a fair share is only somewhat appropriate (Wilcoxon signed-rank test: T(Base): $z = 3.476$, $p = 0.005$; T(Before): $z = 2.595$, $p = 0.0095$; T(NoIncentive): $z = 4.389$, $p < 0.00$; T(Questions): $z = 3.124$, $p = 0.0018$; PTD: $z = 2.664$, $p = 0.0077$). Empirical expectations are mostly

similarly distributed among treatments, where two-thirds (on average 66.67%) of the participants expect others to provide a fair share (see Figure 5a-c for percentages across treatments).

Result 1: *While the majority find a fair share very appropriate, the majority expect others to find a fair share only somewhat appropriate. However, there is no discrepancy between personal normative beliefs and normative expectations of whether providing a fair share is appropriate or not.*

Figure 5: Distribution (Percentage) Regarding a Fair Share across Treatments

3.3. Test of Predictions

3.3.1. Results regarding Prediction P1 – Time of Elicitation

Concerning the time of norm elicitation, we find no significant difference between eliciting personal normative beliefs on the appropriateness of providing a fair share (Option C or D) before in T(Before) or after the decision in T(Base). Applying a one-sided Mann-Whitney U Test reveals that the appropriateness ratings of a fair share are not significantly different ($z = 0.127$, $p = 0.44955$). Our results also reveal that normative expectations measured before and after the decision do not differ significantly (one-sided Mann-Whitney U Test: $z = -1.148$, $p = 0.12545$). However, slightly significantly more participants estimated the correct normative expectation (that a fair share is very appropriate) when eliciting normative expectations before in T(Before) than after the decision (one-sided Chi-Square Test: $\chi^2(1) = 2.182$, $p = 0.07$). This difference in accuracy may be due to participants having a more positive view of the normativity of others before than after having themselves made the decision.

In respect of actual behaviour, results indicate that decisions to provide a fair share are not significantly different when beliefs are elicited before and after the task, leading us to reject H1 (one-sided Chi-Square Test: $\chi^2(1) = 0.5075$, $p = 0.238$). Empirical expectations do not differ either (two-sided Chi-Square Test $\chi^2(1) = 1.0933$, $p = 0.296$).

Result 2: *Behaviour does not significantly differ when eliciting personal normative beliefs and normative expectations before the task compared to after the task, leading us to reject P1.*

3.3.2. Results regarding Prediction P2 – Incentivizing Elicitation

Normative expectations for a fair share are considered significantly more appropriate when they are incentivized in T(Base) compared to when they are not in T(NoIncentive) (one-sided Mann-Whitney U Test: $z = 2.860$ $p = 0.0021$). Referring to the second Prediction (P2), applying a one-sided Chi-Square Test reveals that the incentivization of normative expectations has a significant impact on accuracy ($\chi^2(1) = 3.7400$, $p = 0.0265$). The share of correct fits (normative expectations matching the most common actual response) is higher in T(Base) with incentivization than in T(NoIncentive) without an incentive – supporting Prediction P2. Moreover, we find no significant differences between the two treatments regarding personal normative beliefs (two-sided Mann-Whitney U Test: $z = 0.685$, $p = 4935$), behaviour ($z = -0.237$, $p = 0.1824$), and empirical expectations (two-sided Chi-Square Test $\chi^2(1) = 0.0027$, $p = 0.959$).

Result 3: *Incentivization significantly increases the accuracy of normative expectations, supporting P2.*

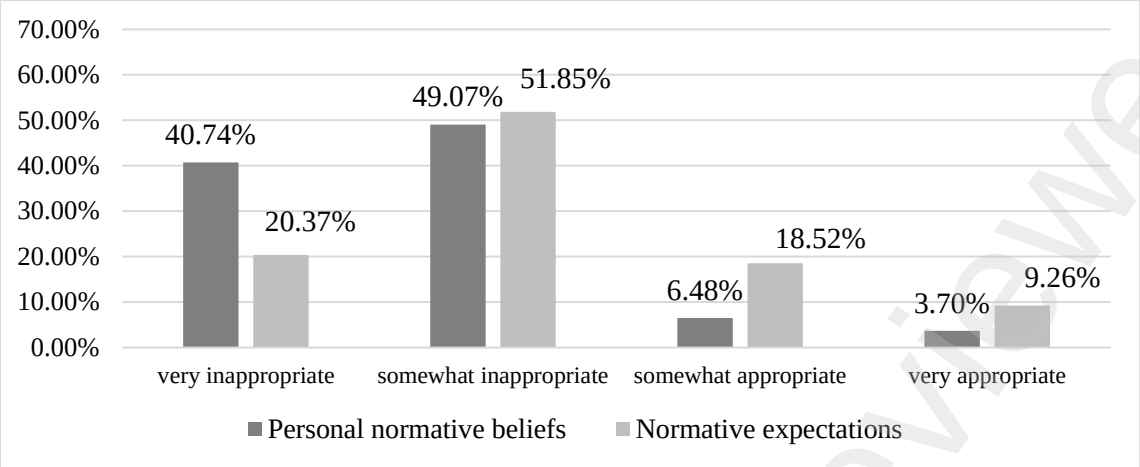
3.3.3. Results regarding Predictions P3.1-P3.3 – Questions about a Fair and Low Share

When asking participants about their belief about not just one behavioural action (a fair share), but at least two behavioural alternatives (fair share and low share) both personal normative beliefs and normative expectations about the appropriateness of a fair sharing decrease. Put differently, when participants were asked to consider their own beliefs not only about whether it's right to provide a fair share, but also whether it is valid to provide a low share, they were less likely to believe in the appropriateness of providing a fair share, or to believe that others shared that view. However, the decrease is neither significant either for personal normative beliefs in P3.1 (two-sided Mann-Whitney U Test: $z = 1.198$, $p = 0.2311$) nor for normative expectations in P3.2 (two-sided Mann-Whitney U Test: $z = 1.226$, $p = 0.2204$). Regarding empirical expectations, our results indicate that significantly fewer participants in T(Base) expect others to provide a fair share, compared to those in T(Questions) (one-sided Chi-Square Test (1) = 3.0754, $p = 0.0395$), leading us to reject Prediction P3.3. Asking personal normative beliefs and normative expectations about both alternatives might have increased the salience of a fair share. Afterall, most participants indicated a personal normative belief and normative expectations of a low share as inappropriate, which might have increased the expectation that others provided a fair share.

Results 4: *Asking about at least two behavioural alternatives instead of only one (the main option of interest) does not significantly influence personal normative beliefs and normative expectations, providing support for P3.1 and P3.2. But it increases empirical expectations about the behaviour of interest, leading us to reject P3.3.*

In the next section, we further analyse participants answers regarding a low share. Concerning answers about a low share, most participants hold personal normative beliefs (89.91%) and normative expectations (72.22%) for a low share to be (somewhat or very) inappropriate. In contrast, only 10.18% of the subjects stated that they find providing a low share (somewhat or very) appropriate, while more than twice as many (27.78%) expect others to believe a low share to be appropriate. About half of the participants held personal normative beliefs and normative expectations of a low share to be somewhat inappropriate (Figure 6). Interestingly, only half as many believe a low share to be very inappropriate (40.74%) and expect others to share this belief (20.37%). By contrast, while only about 10% personally believe a low share to be either somewhat or very appropriate, their number nearly triples for normative expectations (27.78%) (Figure 6). The mean value for normative expectations of a low share in T(Questions) is 2.17 (sd = 0.08), and for personal normative beliefs it is 1.73 (sd = 0.07) – a significant difference from normative expectations (Wilcoxon signed-rank test: $z = -4.499$, $p < 0.00$).

Figure 6: Percentage Distribution of a Low Share in T(Questions)



After all, asking about two behavioural alternatives has the advantage of providing data for additional analyses. For instance, it allows to investigate whether participants have multiple normative expectations (e.g., expecting that others find two opposing actions similarly appropriate or inappropriate), whether the behaviour is subject to polarized norms (e.g., some individuals find one action appropriate and the other inappropriate, while for others the exact opposite is true), or whether an unambiguous norm prevails. In this study, combining both normative expectations, we find that two-thirds (66.67%) of the participants expect a fair share to be appropriate and a low share to be inappropriate (Table 4), indicating a consistent expectation in favour of a fair share for the majority. Still, one-quarter of participants expect both to be appropriate, indicating the presence of multiple normative expectations. This may also result from the fact that normative expectations for a low share were more widely distributed than those for a fair share. Expecting both behaviours to be appropriate might create cognitive dissonance and uncertainty about how to behave. In this case, it might be worth investigating whether providing information about the true majority belief (a low share being inappropriate) might resolve this ambiguity, thereby reducing cognitive dissonance and leading to an even greater increase in empirical expectations and behaviours favouring a fair share.

Table 3: Multiple Normative Expectations

		Normative expectation: fair share		
		Inappropriate	Appropriate	Σ
Normative expectation: low share	Inappropriate	6 (5.56%)	72 (66.67%)	78 (72.22%)
	Appropriate	4 (3.70%)	26 (24.07%)	30 (27.78%)
	Σ	10 (9.26%)	98 (90.74%)	108

Results 5: *In line with the majority norm, the majority has a normative expectation that a fair share is appropriate and a low share inappropriate. However, one-quarter finds both actions appropriate,*

which does not reflect the majority norm, and which might be resolved by providing information on the majority norm concerning a low share.

3.4. Future Research & Limitations

Our study carries three main potentials for future research. First, concerning empirical expectations, participants in this study were only asked whether they expect most others to provide a fair share (dummy variable: yes/no). However, assessing the expected distribution would have been interesting for a more detailed analysis (e.g., "how many other participants, in percentage terms, do you think decided in favour of a fair split?"). The advantage of asking for a concrete distribution would have been to gain more detailed information about the distribution and how variations in the elicitation procedure of normative expectations may affect empirical expectations.

A second potential arises from assessing social norms in a specific cultural context. Social norms and the responses regarding the variation of the norm elicitation procedure may vary between different cultures. This study was conducted in the UK, and while its results concerning the effects of the elicitation method may extend/transfer to other Western cultures, repetition could test these finding in other cultures and regions.

Third, our results only hold for behaviours with a predominantly unambiguous social norm – in our context, fairness. Future research might investigate whether results also apply to behaviours with a more unambiguous or even misperceived social norm. Other social norms might not be as salient as the fairness norm is for the budget division.

4. Conclusion

The two-step norm elicitation method pioneered by Bicchieri and Xiao (2009) provides an easy-to-implement technique that captures personal normative beliefs and normative expectations about behaviours. Asking about the appropriateness of behaviour on a four-point scale provides researchers with more nuanced results, especially for detecting pluralistic ignorance and the misperception of beliefs. This study investigated the sensitivity of variations in eliciting fairness beliefs and their impact on behaviours in a dictator game, using the two-step norm elicitation method. Running an online experiment with 438 subjects, we tested three different treatments. In line with previous studies, our results support the idea that incentives do make a difference. More specifically, incentivizing normative expectations significantly increases their correct estimation of the majority norm. By contrast to previous studies, the results of our experimental design indicate robustness for the time of elicitation: eliciting beliefs before the decision task did not influence behaviour compared to elicitation after the decision. The explanation might be two-fold: first, the rate of participants providing a fair share is already high across all treatments (on average, 66.67%). Second, the social norm of fairness is potentially well-known. Hence, in this study, elicitation before the task may not have had a priming

effect, as the social norm already has been salient. This might be different in a more complex context, or when beliefs and behaviours are elicited about norms that entail more uncertainty. Finally, asking participants about the appropriateness of both a fair share and a low share did not significantly increase personal normative beliefs nor normative expectations. However, it did significantly increase empirical expectations about a fair share. Finally, the elicitation of normative expectations of both, a fair share and a low share, revealed that the fairness norm is very unambiguous, but one-quarter falsely believed that the majority regarded both shares as appropriate. The insights from this study may guide researchers who want to use the two-step norm elicitation method.

Appendix

Table A1: Descriptive Analysis across Treatments

	PTD	PTR	T(Base)	T(Before)	T(NoIncentive)	T(Questions)
Total (n)	102	99	111	109	110	108
Age (mean)	40.08	38.69	42.11	40.58	39.76	39.21
Gender						
Female	61 (59.80%)	56 (56.57%)	58 (52.25%)	52 (47.71%)	54 (49.09%)	66 (61.11%)
Male	41 (20.20)	42 (42.42%)	53 (47.75%)	56 (51.38%)	56 (50.91%)	42 (38.89%)
Non-binary	-	1 (1.01%)	-	1 (0.92%)	-	-
Education						
Student in full-time education	-	1 (1.01%)	2 (1.80%)	1 (0.92%)	1 (0.91%)	-
School leavers without qualification	4 (3.92%)	1 (1.01%)	2 (1.80%)	1 (0.92%)	1 (0.91%)	2 (1.85%)
GCSE Level	18 (17.65%)	12 (12.12%)	15 (13.51%)	15 (13.76%)	14 (12.73%)	12 (11.11%)
Completed apprenticeship	1 (0.98%)	5 (5.05%)	5 (3.60%)	5 (4.59%)	6 (5.45%)	5 (4.63%)
A-Level	20 (19.61%)	22 (22.22%)	11 (9.91%)	21 (19.27%)	12 (10.91%)	20 (18.52%)
Undergraduate degree	41 (40.20%)	37 (37.37%)	52 (46.85%)	45 (41.28%)	52 (47.27%)	50 (46.30%)
Postgraduate degree	18 (17.65%)	21 (21.21%)	20 (18.02%)	16 (14.68%)	22 (20%)	14 (12.96%)
PhD	-	-	5 (4.50%)	5 (4.59%)	2 (1.82%)	5 (4.63%)
Employment status						
Yes, full-time	53 (51.96%)	53 (51.96%)	68 (61.26%)	55 (50.46%)	58(52.73%)	60 (55.56%)
Yes, part-time	24.51%)	25 (24.51%)	22 (19.82%)	25 (22.94%)	30 (27.27%)	24 (22.22%)
No	24 (23.53%)	24 (23.53%)	21 (18.92%)	29 (26.61%)	22 (20%)	24 (22.22%)
Role at work						
Upper management	2 (1.96%)	4 (4.04%)	3 (2.70%)	3 (2.75%)	2 (1.82%)	5 (4.63%)
Trained professional	20 (19.61)	12 (12.12%)	13 (11.71%)	21 (19.27%)	13 (11.82%)	17 (15.74%)
Middle management	15 (14.71%)	19 (19.19%)	22 (19.82%)	16 (14.68%)	19 (17.27%)	17 (15.74%)
Skilled labourer	7 (6.86%)	8 (8.08%)	4 (3.60%)	3 (2.75%)	2 (1.82%)	4 (3.70%)
Junior management	9 (8.82%)	8 (8.08%)	17 (15.32%)	6 (5.50%)	9 (8.18%)	4 (3.70%)
Consultant	2 (1.96%)	5 (5.05%)	-	1 (0.92%)	5 (4.55%)	5 (4.63%)
Administrative staff	10 (9.80%)	13 (13.13%)	16 (14.41%)	15 (13.76%)	17 (15.45%)	15 (13.89%)
Temporary employee	2 (1.96%)	-	1 (0.90%)	1 (0.92%)	-	-
Support staff	5 (4.90%)	3 (3.03%)	6 (5.41%)	7 (6.42%)	5 (4.55%)	3 (2.78%)
Researcher	-	1 (1.01%)	1 (0.90%)	-	1 (0.91%)	1 (0.93%)
Student	3 (2.94%)	3 (3.03%)	2 (1.80%)	2 (1.83%)	5 (4.55%)	3 (2.73%)
Self-employed/Partner	7 (6.86%)	6 (6.06%)	7 (6.31%)	9 (8.26%)	12 (10.01%)	8 (7.41%)
Other	3 (2.94%)	5 (5.05%)	3 (2.70%)	5 (4.59%)	3 (2.73%)	4 (3.70%)
Not working	17 (16.67%)	12 (12.12%)	16 (14.41%)	20 (18.35%)	17 (15.45%)	22 (20.37%)
Decision / Share						
A (£2.25 / £0.25)	22 (21.57%)		18 (16.22%)	16 (14.68%)	17 (15.45%)	15 (13.89%)
B (£2.00 / £0.50)	6 (5.88%)		9 (8.11%)	8 (7.34%)	11 (10%)	7 (6.48%)
C (£1.50 / £1.00)	16 (15.69%)		15 (13.51%)	13 (11.93%)	12 (10.91%)	12 (11.11%)
D (£1.25 / £1.25)	57 (55.88%)		67 (60.36%)	72 (66.06%)	67 (60.91%)	74 (68.52%)
E (£1.00 / £1.50)	1 (0.98%)		1 (0.90%)	-	-	-
F (£0.50 / £2.00)	-		-	-	1 (0.91%)	-
G (£0.25 / £2.25)	-		1 (0.90%)	-	2 (1.82%)	-
PNB re fair share						
Very inappropriate	7 (6.86%)	4 (4.04%)	3 (2.70%)	1 (0.92%)	2 (1.82%)	3 (2.78%)
Somewhat inappropriate	4 (3.92%)	2 (2.02%)	2 (1.80%)	2 (1.83%)	11 (10%)	2 (1.85%)
Somewhat appropriate	30 (29.41%)	29(29.29%)	37 (33.33%)	40 (36.70%)	31 (28.18%)	45 (41.67%)
Very appropriate	61 (59.80%)	64 (64.95%)	69 (62.16%)	66 (60.55%)	66 (60%)	58 (53.70%)
PNB re low share						
Very inappropriate						44 (40.74%)
Somewhat inappropriate						53 (49.07%)
Somewhat appropriate						7 (6.48%)
Very appropriate						4 (3.70%)
NE re fair share						
Very inappropriate	5 (4.90%)	2 (2.02%)	2 (1.80%)	2 (1.83%)	2 (1.82%)	2 (1.85%)
Somewhat inappropriate	3 (2.94%)	7 (7.07%)	2 (1.80%)	5 (4.59%)	17 (15.45%)	8 (7.41%)
Somewhat appropriate	53 (51.96%)	48 (48.48%)	62 (55.86)	47 (43.12%)	60 (54.55%)	60 (54.55%)
Very appropriate	41 (40.20%)	42 (42.42%)	45 (40.54%)	55 (50.46%)	31 (28.18%)	38 (35.19%)

NE re low share						
Very inappropriate						22 (20.37%)
Somewhat inappropriate						56 (51.85%)
Somewhat appropriate						20 (18.52%)
Very appropriate						10 (9.26%)
EE re fair slit						
yes	65 (63.73%)	64 (64.65%)	70 (63.06%)	76 (69.72%)	69 (62.73%)	80 (74.07%)
no	37 (36.27%)	35 (35.35%)	41 (36.94%)	33 (30.28%)	41 (37.27%)	28 (25.93%)

Note: PTD = Pretest Dictator, PTR = Pretest Receiver, PNB = Personal normative beliefs, NE = Normative expectations, EE = Empirical expectations

References

Abeler, J., Nosenzo, D., & Raymond, C. (2019). Preferences for truth-telling. *Econometrica*, *87*(4), 1115–1153.

Agerström, J., Carlsson, R., Nicklasson, L., & Guntell, L. (2016). Using descriptive social norms to increase charitable giving: The power of local norms. *Journal of Economic Psychology*, *52*, 147–153. https://doi.org/10.1016/j.joep.2015.12.007

Albert, D. A., & Smilek, D. (2023). Comparing attentional disengagement between Prolific and MTurk samples. *Scientific Reports*, *13*(1), 20574. https://doi.org/10.1038/s41598-023-46048-5

Andreoni, J., & Bernheim, B. D. (2009). Social Image and the 50-50 Norm: A Theoretical and Experimental Analysis of Audience Effects. *Econometrica*, *77*(5), 1607–1636. https://doi.org/10.3982/ECTA7384

Aycinena, D., Bogliacino, F., & Kimbrough, E. O. (2024). Measuring norms: Assessing the threat of social desirability bias to the Bicchieri and Xiao elicitation method. *Journal of Economic Behavior & Organization*, *222*, 225–239. https://doi.org/10.1016/j.jebo.2024.04.002

Barr, A., Lane, T., & Nosenzo, D. (2018). On the social inappropriateness of discrimination. *Journal of Public Economics*, *164*, 153–164. https://doi.org/10.1016/j.jpubeco.2018.06.004

Bicchieri, C. (2006). *The grammar of society: The nature and dynamics of social norms*. Cambridge University Press.

Bicchieri, C. (2017). *Norms in the wild: How to diagnose, measure, and change social norms*. Oxford University Press.

Bicchieri, C., & Chavez, A. (2010). Behaving as expected: Public information and fairness norms. *Journal of Behavioral Decision Making*, *23*(2), 161–178. https://doi.org/10.1002/bdm.648

Bicchieri, C., Dimant, E., Gachter, S., & Nosenzo, D. (2020). Observability, Social Proximity, and the Erosion of Norm Compliance. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3576289

Bicchieri, C., Dimant, E., & Sonderegger, S. (2023). It's not a lie if you believe the norm does not apply: Conditional norm-following and belief distortion. *Games and Economic Behavior*, *138*, 321–354. https://doi.org/10.1016/j.geb.2023.01.005

Bicchieri, C., Dimant, E., & Xiao, E. (2021). Deviant or wrong? The effects of norm information on the efficacy of punishment. *Journal of Economic Behavior & Organization*, *188*, 209–235.

Bicchieri, C., Muldoon, R., & Sontuoso, A. (2018). *Social norms*.

Bicchieri, C., & Xiao, E. (2009). Do the right thing: But only if others do so. *Journal of Behavioral Decision Making*, *22*(2), 191–208. https://doi.org/10.1002/bdm.621

Bogliacino, F., Charris, R., Gómez, C., & Montealegre, F. (2024). Negative economic shocks and the compliance to social norms. *Judgment and Decision Making*, *19*, e9. https://doi.org/10.1017/jdm.2024.1

Bolton, G. E., Katok, E., & Zwick, R. (1998). Dictator game giving: Rules of fairness versus acts of kindness. *International Journal of Game Theory*, *27*(2), 269–299. https://doi.org/10.1007/s001820050072

Brañas-Garza, P. (2007). Promoting helping behavior with framing in dictator games. *Journal of Economic Psychology*, *28*(4), 477–486. https://doi.org/10.1016/j.joep.2006.10.001

Bursztyn, L., Egorov, G., & Fiorin, S. (2020). From Extreme to Mainstream: The Erosion of Social Norms. *American Economic Review*, *110*(11), 3522–3548. https://doi.org/10.1257/aer.20171175

Bursztyn, L., González, A. L., & Yanagizawa-Drott, D. (2020). Misperceived Social Norms: Women Working Outside the Home in Saudi Arabia. *American Economic Review*, *110*(10), 2997–3029. https://doi.org/10.1257/aer.20180975

Camerer, C. F. (2003). *Behavioral game theory: Experiments in strategic interaction*. Princeton University Press.

Camerer, C. F., & Thaler, R. H. (1995). Anomalies: Ultimatums, Dictators and Manners. *Journal of Economic Perspectives*, *9*(2), 209–219. https://doi.org/10.1257/jep.9.2.209

Castillo, G., Choo, L., & Grimm, V. (2022). *Do different people report the same social norms?*

Chen, D. L., Schonger, M., & Wickens, C. (2016). oTree—An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, *9*, 88–97. https://doi.org/10.1016/j.jbef.2015.12.001

Cialdini, R. B., & Goldstein, N. J. (2004). Social Influence: Compliance and Conformity. *Annual Review of Psychology*, *55*(1), 591–621. https://doi.org/10.1146/annurev.psych.55.090902.142015

Cox, J. C. (2010). Some issues of methods, theories, and experimental designs. *Journal of Economic Behavior & Organization*, *73*(1), 24–28. https://doi.org/10.1016/j.jebo.2009.01.014

D'Adda, G., Drouvelis, M., & Nosenzo, D. (2016). Norm elicitation in within-subject designs: Testing for order effects. *Journal of Behavioral and Experimental Economics*, *62*, 1–7. https://doi.org/10.1016/j.socec.2016.02.003

D'Adda, G., Dufwenberg, M., Passarelli, F., & Tabellini, G. (2020). Social norms with private values: Theory and experiments. *Games and Economic Behavior*, *124*, 288–304. https://doi.org/10.1016/j.geb.2020.08.012

Dreber, A., Ellingsen, T., Johannesson, M., & Rand, D. G. (2013). Do people care about social context? Framing effects in dictator games. *Experimental Economics*, *16*(3), 349–371. https://doi.org/10.1007/s10683-012-9341-9

Engel, C. (2011). Dictator games: A meta study. *Experimental Economics*, *14*(4), 583–610. https://doi.org/10.1007/s10683-011-9283-7

Epley, N., & Gilovich, T. (2005). When effortful thinking influences judgmental anchoring: Differential effects of forewarning and incentives on self-generated and externally provided anchors. *Journal of Behavioral Decision Making*, *18*(3), 199–212. https://doi.org/10.1002/bdm.495

Fallucchi, F., & Nosenzo, D. (2022). The coordinating power of social norms. *Experimental Economics*, *25*(1), 1–25. https://doi.org/10.1007/s10683-021-09717-8

Fehr, E., & Fischbacher, U. (2004). Third-party punishment and social norms. *Evolution and Human Behavior*, *25*(2), 63–87.

Fehr, E., & Schmidt, K. M. (2006). Chapter 8 The Economics of Fairness, Reciprocity and Altruism – Experimental Evidence and New Theories. In *Handbook of the Economics of Giving, Altruism and Reciprocity* (Bd. 1, S. 615–691). Elsevier. https://doi.org/10.1016/S1574-0714(06)01008-6

Gächter, S., Gerhards, L., & Nosenzo, D. (2017). The importance of peers for compliance with norms of fair sharing. *European Economic Review*, *97*, 72–86. https://doi.org/10.1016/j.euroecorev.2017.06.001

Gächter, S., & Renner, E. (2010). The effects of (incentivized) belief elicitation in public goods experiments. *Experimental Economics*, *13*(3), 364–377. https://doi.org/10.1007/s10683-010-9246-4

Gneezy, U., Saccardo, S., & Van Veldhuizen, R. (2019). Bribery: Behavioral Drivers of Distorted Decisions. *Journal of the European Economic Association*, *17*(3), 917–946. https://doi.org/10.1093/jeea/jvy043

Görges, L., & Nosenzo, D. (2020). Measuring Social Norms in Economics: Why It Is Important and How It Is Done. *Analyse & Kritik*, *42*(2), 285–312. https://doi.org/10.1515/auk-2020-0012

Graham, J., Nosek, B. A., Haidt, J., Iyer, R., Koleva, S., & Ditto, P. H. (2011). Mapping the moral domain. *Journal of Personality and Social Psychology*, *101*(2), 366–385. https://doi.org/10.1037/a0021847

Heinrich, T., & Weimann, J. (2013). A note on reciprocity and modified dictator games. *Economics Letters*, *121*(2), 202–205. https://doi.org/10.1016/j.econlet.2013.08.004

Horne, C., & Mollborn, S. (2020). Norms: An Integrated Framework. *Annual Review of Sociology*, *46*(1), 467–487. https://doi.org/10.1146/annurev-soc-121919-054658

Iriberri, N., & Rey-Biel, P. (2011). The role of role uncertainty in modified dictator games. *Experimental Economics*, *14*(2), 160–180. https://doi.org/10.1007/s10683-010-9261-5

Kassas, B., & Palma, M. A. (2019). Self-serving biases in social norm compliance. *Journal of Economic Behavior & Organization*, *159*, 388–408. https://doi.org/10.1016/j.jebo.2019.02.010

Kimbrough, E. O., & Vostroknutov, A. (2016). NORMS MAKE PREFERENCES SOCIAL: Norms Make Preferences Social. *Journal of the European Economic Association*, *14*(3), 608–638. https://doi.org/10.1111/jeea.12152

Kimbrough, E. O., & Vostroknutov, A. (2018). A portable method of eliciting respect for social norms. *Economics Letters*, *168*, 147–150. https://doi.org/10.1016/j.econlet.2018.04.030

König-Kersting, C. (2024). On the robustness of social norm elicitation. *Journal of the Economic Science Association*. https://doi.org/10.1007/s40881-024-00178-2

Krupka, E. L., & Weber, R. A. (2013). Identifying Social Norms Using Coordination Games: Why does Dictator Game Sharing Vary? *Journal of the European Economic Association*, *11*(3), 495–524. https://doi.org/10.1111/jeea.12006

Lane, T., Nosenzo, D., & Sonderegger, S. (2023). Law and Norms: Empirical Evidence. *American Economic Review*, *113*(5), 1255–1293. https://doi.org/10.1257/aer.20210970

Mesa-Vázquez, E., Rodriguez-Lara, I., & Urbano, A. (2021). Standard vs random dictator games: On the effects of role uncertainty and framing on generosity. *Economics Letters*, *206*, 109981. https://doi.org/10.1016/j.econlet.2021.109981

Mir Djawadi, B., Plaß, S., & Schäfers, S. (2025). *"I don't believe that you believe what I believe": An experiment on misperceptions of social norms and whistleblowing*. https://dx.doi.org/10.2139/ssrn.4868252

Palan, S., & Schitter, C. (2018). Prolific.ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, *17*, 22–27. https://doi.org/10.1016/j.jbef.2017.12.004

Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology*, *70*, 153–163. https://doi.org/10.1016/j.jesp.2017.01.006

Reuben, E., & Riedl, A. (2013). Enforcement of contribution norms in public good games with heterogeneous populations. *Games and Economic Behavior*, *77*(1), 122–137. https://doi.org/10.1016/j.geb.2012.10.001

Sargent, R. H., & Newman, L. S. (2021). Pluralistic Ignorance Research in Psychology: A Scoping Review of Topic and Method Variation and Directions for Future Research. *Review of General Psychology*, *25*(2), 163–184.

Selten, R. (1967). Die Strategiemethode zur Erforschung des eingeschränkt rationale Verhaltens im Rahmen eines Oligopolexperiments. *Beiträge zur experimentellen Wirtschaftsforschung*.

Veselý, Š. (2015). Elicitation of normative and fairness judgments: Do incentives matter? *Judgment and Decision Making*, *10*(2), 191–197. https://doi.org/10.1017/S1930297500003958

Walkowitz, G. (2021). Dictator game variants with probabilistic (and cost-saving) payoffs: A systematic test. *Journal of Economic Psychology*, *85*, 102387. https://doi.org/10.1016/j.joep.2021.102387