

Charness, Gary; Dimant, Eugen; Gneezy, Uri; Krupka, Erin

**Working Paper**

## Experimental Methods: Eliciting and Measuring Social Norms

CESifo Working Paper, No. 11840

**Provided in Cooperation with:**

Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

*Suggested Citation:* Charness, Gary; Dimant, Eugen; Gneezy, Uri; Krupka, Erin (2025) : Experimental Methods: Eliciting and Measuring Social Norms, CESifo Working Paper, No. 11840, CESifo GmbH, Munich

This Version is available at:

<https://hdl.handle.net/10419/319208>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

## Experimental Methods: Eliciting and Measuring Social Norms

*Gary Charness, Eugen Dimant, Uri Gneezy, Erin Krupka*

## **Impressum:**

CESifo Working Papers

ISSN 2364-1428 (electronic version)

Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH

The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute

Poschingerstr. 5, 81679 Munich, Germany

Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email [office@cesifo.de](mailto:office@cesifo.de)

Editor: Clemens Fuest

<https://www.cesifo.org/en/wp>

An electronic version of the paper may be downloaded

- from the SSRN website: [www.SSRN.com](http://www.SSRN.com)
- from the RePEc website: [www.RePEc.org](http://www.RePEc.org)
- from the CESifo website: <https://www.cesifo.org/en/wp>

# Experimental Methods: Eliciting and Measuring Social Norms

Gary Charness<sup>a</sup>, Eugen Dimant<sup>b,c</sup>, Uri Gneezy<sup>d</sup>, Erin Krupka<sup>e</sup>

<sup>a</sup>*University of California, Santa Barbara, USA*

<sup>b</sup>*University of Pennsylvania, USA*

<sup>c</sup>*CESifo, Germany*

<sup>d</sup>*University of California, San Diego, USA*

<sup>e</sup>*University of Michigan, USA*

## Abstract

Eliciting social norms is essential for understanding a range of behaviors in economic contexts. This paper reviews key experimental approaches to social-norm measurement, comparing the methods, practical considerations, and specific conditions under which each is most effective. We discuss various social norm elicitation techniques, including coordination-based, opinion-based, and distributional approaches. Our findings suggest that coordination-game approaches are the most widely adopted and tested; they deliver robust results, particularly in contexts with a single dominant norm. Importantly, while early methods focused on eliciting mean or modal normative beliefs, recent work shifts focus to eliciting beliefs about the distribution of normative beliefs. This allows the researcher to draw inferences on the degree of uncertainty that underlies norm assessments. This paper aims to help experimentalists and practitioners choose suitable norm-elicitation methods that are aligned with research objectives and logistical constraints.

**Keywords:** Methodology, Norm Elicitation, Social Norms

**JEL Codes:** C91, D91, Z13

---

This paper was conceptualized before Gary's passing. He contributed to earlier iterations but was not involved in the final one. We deeply felt his absence during the revision process, as we do in so many other aspects of life. We are grateful to Giulia Andrighetto, Diego Aycinena, Zvonimir Bašić, Cristina Bicchieri, Erik Kimbrough, Christian König, Daniele Nosenzo, Folco Panizza, Andrea Robbett, Eugenio Verrina, Roberto Weber, and Erte Xiao for their helpful comments and suggestions. Ready-made scripts for the norm elicitation methods discussed in this paper are available for download via our OSF repository: [link](#).

*Email:* [edimant@sas.upenn.edu](mailto:edimant@sas.upenn.edu) (Eugen Dimant, corresponding author), [ugneezy@ucsd.edu](mailto:ugneezy@ucsd.edu) (Uri Gneezy), [ekrupka@umich.edu](mailto:ekrupka@umich.edu) (Erin Krupka), Sadly, Gary Charness passed away during the preparation of this manuscript.

# 1 Introduction

Economic activity occurs in social contexts, which structure and co-determine preferences, beliefs, and constraints. To predict basic activities such as cooperation, coordination, and strategic interaction, one therefore needs answers to normative questions like “How ought I to behave in this situation?” or “How should others behave in this situation?” Research shows that these questions cannot be answered without a rich and contextual knowledge of the social landscape in which the actors find themselves. Social norms are critical features in that landscape (Axelrod, 1986; Elster, 1989, 1991; Coleman, 1994).

Social norms are of interest to economists because they offer a possible explanation for behaviors that otherwise appear inconsistent with standard assumptions, such as self-interest. For example, they can help explain why people share (Krupka and Weber, 2013), conform to others (Bicchieri et al., 2022; Dimant et al., 2024a), are honest (Gneezy, 2005; Abeler et al., 2019), cooperate (Reuben and Riedl, 2013), and are willing to punish those who do not cooperate (Güth et al., 1982; Fehr and Gächter, 2000a; Dimant and Gesche, 2023). They can also help us understand why some communities can solve commons problems while others cannot (Hardin, 1982; Ostrom, 2000), why firms adopt particular price-setting behaviors and do not cut wages during periods of high unemployment (Akerlof, 1980; Kahneman et al., 1986), and why countries adopt different redistribution policies (Alesina and Angeletos, 2005; Almås et al., 2020). They also allow agents to coordinate actions and expectations in interactions with multiple equilibria (Sugden, 1998; Binmore and Samuelson, 2006; Bicchieri et al., 2023).

Research on the examination and quantification of norms, as objects of study in their own right, is growing in importance (Bicchieri and Xiao, 2009; Bicchieri and Chavez, 2010; Krupka and Weber, 2013; Kimbrough and Vostroknutov, 2016; Bicchieri and Dimant, 2019; Dimant, 2023; Aycinena et al., 2024a; Gelfand et al., 2024; Panizza et al., 2024b). For this reason, our paper focuses on the major approaches in experimental economics to eliciting social norms, which vary across multiple dimensions, including what they measure, how they measure it, and the situations in which they can be applied. In particular, we examine the recent developments of social norm measurements, particularly the novel techniques that have been proposed over the past two decades. We describe the most frequently used methods, discuss their strengths and weaknesses, compare them, and preview new developments. We aim to provide a reference and guide for social scientists and practitioners interested in applying norm elicitation methods.

## 2 The Definition of a Social Norm

Though definitions of social norms vary across papers and fields, they share a few elements. Most definitions of norms contain common knowledge or joint recognition among a group that invokes the idea that a collectively believed standard, ideal, or regularity of behavior exists for a situation (Bicchieri, 2006; Krupka and Weber, 2013). For example, Fehr and Schurtenberger (2018) define norms as “commonly known standards of behavior that are based on widely shared views of how individual group members ought to behave in a given situation.” Nosenzo and Görges (2020) define norms as “shared understandings within a social group about what is considered acceptable

or unacceptable behavior in a given situation.” Fehr and Gächter (2000b) write that a norm is defined as “a behavioral regularity; that is [...] based on a socially shared belief about how one ought to behave,” and Ostrom (2000) defines norms as “shared understandings about actions that are obligatory, permitted, or forbidden.” However, Bicchieri (2006) notes that these definitions leave unaddressed the condition by which such shared understandings are distinguished from wide-spread personal normative beliefs and treats the assumption of conditional preferences to adhere to norms in her work; Coleman (1994) foregrounds internal and external sanction avoidance which could also provide a way to distinguish a social norm from personal normative beliefs. We will adopt a working definition:

*“Injunctive social norms are commonly believed standards for behavior among members of a social group in a given situation.”*

Definitions of norms explicitly or implicitly distinguish between social, personal, and descriptive norms, as well as taboos. Schwartz and Fleishman (1982) define personal norms as “self-expectations for behavior constructed in specific situations on the basis of generalized internalized values” (for similar distinctions, see also Elster, 1989; Bicchieri, 2006; Bicchieri and Sontuoso, 2020; Burks and Krupka, 2012; Bašić and Verrina, 2024). Cialdini et al. (1990) define descriptive norms as prevalent or common behavior (see also Bicchieri, 2006; Sliwka, 2007; Bicchieri and Xiao, 2009; Krupka and Weber, 2009; Bursztyn et al., 2020a; Charness et al., 2021; Heinicke et al., 2022, for a treatment of empirical expectations). Definitions of social norms that include reference to descriptive norm beliefs or empirical expectations are usually not assuming that norm compliance generates utility and instead consider compliance the result of conditional preferences for conformity (see Elster, 1989; Bicchieri, 2006; Young, 2015; Bicchieri and Sontuoso, 2020, for discussions on this). A different class of models incorporates beliefs in a self- and other-signaling framework (Benabou and Tirole, 2006; Andreoni and Bernheim, 2009). Individuals care about their reputations and believe that others make inferences about their type (e.g., purely self-serving). Fershtman et al. (2011) distinguish between norms and taboos, noting that norms penalize actions while taboos penalize forbidden thoughts.

Economists have also addressed the relationship between social preferences and norms. One view is that prosocial behavior arises from preferences for prosocial distributions of payoffs (a canonical example of this is, Fehr and Schmidt, 1999). Others model prosocial behavior as a preference to comply with social norms, many of which are pro-social (Akerlof and Kranton, 2000; Kimbrough and Vostroknutov, 2016; Bicchieri et al., 2018). Yet others adopt a framework in which preferences for pro-social behavior are at the individual level, while norms, which are group-level phenomena, function as external constraints on behavior (Fehr and Charness, 2025). More recently, Carpenter and Robbett (2024) provide evidence that both preferences over payoff distributions and adherence to social norms matter for prosocial behavior, with different types of people caring more about outcomes than norm adherence. They concluded that failing to account for norms causes researchers to overestimate the role of outcome-based preferences.

### 3 Established Experimental Approaches to Measuring Social Norms

Measuring injunctive normative beliefs involves assessing social constructs that cannot be measured using non-social instruments. In economics, this measurement is informed by historical approaches grounded in psychology and sociology that rely on the use of surveys, vignettes, or asking people how much they expect others to approve or disapprove of behavior, or on the design of experiments that test for behavior consistent with norms (See for example, [Cialdini et al., 1990](#); [Horne and Mollborn, 2020](#)). Economics researchers have adapted these approaches while seeking to create incentive-compatible methods for identifying social norms.

#### 3.1 Sequential Opinion-Matching

[Bicchieri and Xiao \(2009\)](#) introduce a sequential opinion-matching task that involves two steps. In the first, participants learn about a Dictator Game and are asked if they believe that the dictator should split the money approximately equally (the response format was “yes/no”). In other words, they are asked about their personal normative beliefs about a single action. In the second step, they are asked to guess how many others said “yes” to the first question. In other words, they are asked for their normative expectations about others’ responses. The responses in step one are not incentivized, but in step two, they are; participants are paid if the answer to the second question matches the actual number of “yes” answers.

Subsequent studies asked participants to make a binary evaluation (*fair/not fair, should/should not do*) for more than just one action. In [Bicchieri and Chavez \(2010\)](#) responders were asked to rate three different actions and could indicate if they thought any actions were “fair” (they could choose none, one, or more than one option). In the second step, they were asked to guess the number of participants who thought each action was fair and received a payment if their guess was accurate. [Bicchieri et al. \(2021\)](#) systematically varied the combination of norm information and the (non-)existence of punishment and subsequently measured behavior in a Trust Game. [Lane et al. \(2023\)](#) modified the task to include four response categories (e.g., “How appropriate do you think it is for the man to have sex with the girl? The four response categories are: Very appropriate; Somewhat appropriate; Somewhat inappropriate; Very inappropriate.”).

The response format in [Bicchieri and Xiao \(2009\)](#) asks subjects whether a particular action has the normative characteristic that it “should” be done. A feature of absolute evaluation is that the action is judged independent of the other actions and allows for the expression of positive or negative valence. Step 1 allows the researcher to measure the share of people who rate the action as proscriptive. The incentives in step 2 are such that it should elicit the respondent’s beliefs about the central (modal) tendency of personal normative expectations in a population.

This method focuses on these two beliefs (personal and normative expectations) because they help identify whether a “rule” qualifies as a norm. As [Bicchieri \(2006\)](#) explains, a rule becomes a norm for a population when: i knows that a rule N exists and applies to S-type situations (this is the contingency requirement), i prefers to conform to N in S-type situations on condition that i believes that a sufficiently large subset of P conforms to N in S-type situations (Empirical Expectations), and i believes that a sufficiently large subset of P expects i to conform to N in S-type situations (Normative Expectations), and i believes that a sufficiently large subset of P

expects  $i$  to conform to  $N$  in  $S$ -type situations, prefers  $i$  to conform, and may sanction behavior (Normative Expectations with sanctions). Thus, the approach tests for the belief components necessary to classify an action as a norm.

This approach has several strengths. First, it is easy for subjects to understand. At its core, it involves describing a situation, asking respondents whether they think someone should take a specific action (identified by the experimenter), and then asking them to guess what most others said in response to the first question. This captures both the distribution of personal normative beliefs and the beliefs about the distribution of personal normative beliefs of the same individual, which can be particularly useful for research examining the gap between personal norms, on the one hand, and beliefs about others' personal normative beliefs, on the other, especially when these differences predict behavior. Second, the approach allows researchers to measure the proportion of participants who hold particular second-order normative beliefs about personal normative beliefs in a population.

However, because this approach involves two steps, it can take time to administer, and responding in step 2 requires some sophistication on the part of the participants. The binary evaluation (e.g., should do / should not do) of a single action risks yielding coarse attributes of the norm or obscures the presence of multiple acceptable actions.<sup>1</sup> It is also of note that the question format is not fixed. In some cases, participants are asked what is “fair”; in other cases, what people “should do”; and in still others, “what is appropriate”. The attribute in question (e.g., fair or appropriate) will, in turn, shape participants' responses and would drive the specific norm the researcher ends up identifying.<sup>2</sup> Because the question format depends on the purpose of the research questions and the choice context, this would be a useful tool when there is a specific norm *ex-ante* identified by the researcher but would be less well suited when the objective is to identify *which* normative beliefs are present or to identify whether multiple normative beliefs are present since the question asks respondents to identify the most likely singular normative belief (see [Kimbrough et al., 2024](#), for evidence of multiple normative beliefs even in the dictator game).

Choosing a singleton normative belief may also induce a false-positive identification of norms. [Aycinena et al. \(2024a\)](#) state that a false-positive occurs when the method yields reports of a shared normative expectation when all actions in the set are normatively neutral. Thus, without care or some modifications to the task, the approach might lead either to misinterpretations about the norm, how much agreement there is about the fairest action, and/or a false-positive effect. In addition, in the presence of pluralistic ignorance, first-order personal normative beliefs and second-order beliefs about personal normative beliefs may not coincide. [Bicchieri and Xiao \(2009\)](#), [Bursztyn et al. \(2020b\)](#), and [Nosenzo and Grges \(2020\)](#) describe examples of when this would arise. This wedge is of interest in its own right, but in the context of eliciting social norms,

<sup>1</sup>[Lane et al. \(2023\)](#) offer a promising approach that may remedy these concerns because they expand the binary rating to a four-point scale that ranges from “appropriate” to “inappropriate”.

<sup>2</sup>For example, as [Reuben and Riedl \(2013\)](#) show, there may be both an equality and an equity norm present. However, if using the sequential opinion matching method, participants are prompted to consider what is “fair”, then this might prompt them to consider the equity norm even if they are aware that both norms may legitimately apply to the situation ([Dimant et al., 2024c](#)).



the presence of pluralistic ignorance would not surface an accurate norm in the second stage of this method. Related, if one takes the position that norms are beliefs about what others in a group approve or disapprove of, then one’s own personal normative views regarding what is or is not appropriate are not necessarily relevant (Nosenzo and G3rges, 2020, discuss this). There are also cases where even in the absence of pluralistic ignorance, personal normative beliefs differ from the norm. For example, this could arise if you are an employee at an organization (see Burks and Krupka, 2012; Choi et al., 2017; Bae et al., 2025). Finally, social desirability bias could also cause this approach to yield an inaccurate distribution of personal beliefs and/or guesses about that distribution if the topics being studied are sensitive and respondents want to manage their social image or self-image (see Bursztyn et al., 2020b; Aycinena et al., 2024b). Aycinena et al. (2024b) conclude that social desirability bias is a potential threat when certain preconditions are satisfied: subjects are aware that incentives (image concerns, fear of retaliation, monetary, etc.) exist to misreport their personal normative beliefs in the first stage, and are sophisticated enough to recognize in the second step that others also faced those incentives.

### 3.2 The Coordination-Game Approach

Krupka and Weber (2013) develop an incentivized method to estimate injunctive social norms. The method uses coordination games to generate an empirical proxy for the norm that relies on the assumption that norms create focal points in the coordination game. In this method, participants read a vignette that describes a decision-making scenario and the set of possible actions an actor could take in that setting. They then play a coordination game that involves “pure matching” (Schelling, 1980; Mehta et al., 1994), in which they rate whether actions are socially appropriate or inappropriate. The game incentivizes them to match their ratings with those of others by paying them if their ratings match the modal rating of others in the reference group (often other subjects completing the task). The aggregated responses yield a profile of perceived social appropriateness for each action within the given situation and reference group. This is then treated as an empirical proxy for the social norm.

The response format used in this approach elicits an absolute evaluation of each action as participants rate an action on a four-point ordinal scale. The scale allows for positive and negative valence to be expressed, and it allows empirical characterization of both what most respondents think is most appropriate and of the spread of normative ratings over actions. One way to think about this is that the researcher obtains information on the extensive and intensive margins of second order beliefs: the method offers estimates of the appropriateness of any action in the set and yields information on the degree to which there is strong or weak agreement on the appropriateness rating for that action. The incentives are such that they elicit beliefs about the modal tendency in appropriateness rating.

Three methodological points should be noted. First, a critical difference between the coordination game approach and sequential matching is that the presence of a norm is assumed to create a focal point for each coordination game. Because this is a pure matching game with all equilibria equally likely, the authors argue that it makes shared views of appropriateness the most likely focal point, even under risk neutrality. This is a subtle but fundamental difference

between the coordination approach and many others. The incentives motivate participants to identify focal points in the “rating game” and the instructions nudge subjects to look for particular focal points that stand out in a “normative sense”. This means the coordination-game approach as proposed by Krupka and Weber encourages people to identify (and reveal through their ratings) their belief about *the most focal normative beliefs* in the situation.<sup>3</sup> This distinction fundamentally drives some of the experimental design features, such as how the coordination task is explained. Second, it drives the decision not to have a “neutral” response feature. Krupka and Weber (2013) chose to present participants with four appropriateness categories (*very socially inappropriate, socially inappropriate, socially appropriate, and very socially appropriate*), with no neutral category. They considered it essential to exclude a “neutral” category, as this could create an alternative focal point distinct from the one related to the social norm. They elected four categories on the theory that fewer might hinder participants’ ability to differentiate degrees of appropriateness, whereas more might make it difficult for them to align with the social norm and lead them to rely on other focal points instead. Later studies, however, have used six categories, with the argument that this better captured the nuances in how people evaluate actions that fall between the two most extreme options in the studied action space (Krupka et al., 2017; Chang et al., 2019) or even a continuous scale (Merguei et al., 2022).<sup>4</sup> Finally, it is worth noting that other studies have paid participants using a comparison of their ratings to a randomly selected person rather than the modal rating (Chang et al., 2019). This is theoretically equivalent to using the modal rating for that action but may be easier to explain or implement. Krupka et al. (2017) used a proper scoring rule as an incentive. However, this method is only compatible with restrictive assumptions about participants’ risk attitudes, and it can lead to lower payoffs if participants make extreme judgments.

The data gathered in a coordination game can be used to estimate the weight (sometimes authors will refer to this as “the influence”) that a norm has in decision-making. When the norm ratings are used within a conditional logit regression, where the dependent variable is the selected action and the independent variables are the characteristics of the potential actions (e.g., each action’s social-appropriateness rating and its expected monetary payoff), then the ratio of the coefficients can reveal how much money an individual is willing to forgo to achieve one category higher in social appropriateness. This allows for both estimating the weight of norms in decision-making and for comparisons of the weights across different games or populations (Krupka and Weber, 2013; Krupka et al., 2017; Chang et al., 2019). However, researchers should be cautious of potential scaling issues. Specifically, when comparing experiments using this ratio, variations in the size of the stakes may affect how participants interpret the difference between “very socially appropriate” and “very socially inappropriate,” altering the outcomes’ comparability.

The coordination-game approach has several strengths. As noted in our definition, norms are group-level constructs that reflect collective perceptions. Coordination games present a useful, incentivized way of identifying such constructs. From a game-theoretic point of view, pure

---

<sup>3</sup>We thank Daniele Nosenzo for helping to articulate this point.

<sup>4</sup>Note that for the incentive scheme in the continuous measure is somewhat complex to describe to subjects; further testing would be useful to determine how robust this new approach would be to subject comprehension.

matching games may have multiple equilibria, and nothing intrinsic to the payoffs makes one equilibrium favored (or focal) over the others. This property of the game allows the norm to serve as a focal point. Second, the data yield *norm profiles*, which show how socially appropriate each action is based on ratings of the options available to a decision-maker, but also reveal the relative differences in appropriateness among the actions. Both the most appropriate action and the relative differences in ratings are critical for predicting behavior (the relative differences characterize how costly deviation is). Third, this method integrates easily with simple models of norm-dependent utility, where individuals weigh the utility of outcomes against the utility of complying with norms (Akerlof and Kranton, 2000; Krupka and Weber, 2013). An additional strength is that the method does not rely on direct questions about one’s own beliefs or the appropriateness of one’s own actions, which have been shown to suffer from bias (Nosenzo and Gorges, 2020). The method has been adapted to elicit norms for specific reference groups or social identities (Burks and Krupka, 2012; Chang et al., 2019). And finally, Schmidt et al. (2022) note that the coordination-game approach is useful when eliciting beliefs about items that have no “ground truth” (e.g., hypothetical, counterfactual, or far-future events) while still providing monetary incentives to induce cognitive effort.

The coordination-game approach assumes that local norms are commonly known and widely shared. It would work less well if there were no norms, if there were several norms that could apply to the situation (discussed below in the section on comparison of methods), or if there were strategic coordination. It also provides incentives for subjects to guess what others guess, which may cause them to rely on higher-order beliefs (Huffman et al., 2024); indeed, looking at Krupka and Weber (2013), one can see that many do not successfully match ratings. Lane et al. (2023) discussed this in footnote 16 of their paper and pointed out that a “counterargument is that second-order beliefs are the most salient high-order belief in the task because the instructions are heavily framed in the language of social norms, and, if subjects use salient focal points to coordinate (Schelling, 1980), they should indeed report second-order beliefs.” With that in mind, experimenters need to be careful to set “large enough” incentives for successful coordination. As yet under-explored is what incentive size is sufficient in a particular context. Veselý (2015) found that participants provide virtually the same responses in incentivized and non-incentivized versions of the coordination game for dictator games, but other contexts have not been explored much. A recent working paper by Fallucchi et al. (2025) explores incentives in novel contexts (drug prices, lost wallet, beggar vignette) and also finds no evidence of distortion in responses when incentives are or are not used in the coordination game.

A few other concerns are worth keeping in mind. In using appropriateness categories, the coordination-game approach imposes ratio-scale characteristics on ordinal measurements. In practice, this is for convenience and does not alter inference for some types of analyses, as when using rank-order tests for the equality of distributions. But on other occasions, it introduces implicit assumptions (e.g., linearity) that subsequent analyses must rely on. Additionally, using between one and four (or six) categories implies, by construction, lower standard deviations for random variables with means near the upper or lower bound. Hence, one can observe lower

variance caused by not only the presence of a norm but also the construction of the measure.<sup>5</sup> Finally, explaining the game itself may be complex, adding noise to any inference about norms. Huffman et al. (2024) found evidence suggesting that participants got confused as they played, especially as measures became more complex and strategically uncertain.

### 3.3 Comparisons of Coordination Game and Sequential Opinion Matching Norms Measures

#### 3.3.1 Robustness

Most studies testing the robustness of the two approaches have focused on the coordination-game approach developed by Krupka and Weber (2013). Huffman et al. (2024) do so with online participants, replicating the dictator games referred to as “bully” and “standard,” and confirmed the main results (see also König-Kersting, 2024, who use MTurk participants). Fallucchi and Nosenzo (2022) replicate the standard dictator game and test how sensitive the task is to competing focal points, such as attaching visual labels (smileys or frowny faces) to the social appropriateness ratings that subjects use in the coordination task for a dictator game. They find that the method is robust when there are clear norms for what constitutes appropriate behavior. However, when the norms are not clear (i.e., there is uncertainty), they find some evidence of distortion stemming from a competing focal point (see the working paper by Schmidt, 2019, who also addresses multiple focal points). However, Fallucchi and Nosenzo (2022) were not able to replicate it in a second study, and both Krupka et al. (2017) and Lane et al. (2023) also do not find evidence of competing focal points.

What may explain the different results is that Fallucchi and Nosenzo (2022) used a visual competing focal point, while both Krupka et al. (2017) and Lane et al. (2023) used natural competing focal points (such as a well-known law that is relevant to the context or information about what others are doing). In a recent working paper, Aycinena et al. (2024c) reject the hypothesis that the neutral response category is used as a focal point and provide evidence that it may be used by respondents to authentically reflect neutrality. They too test a treatment that introduces an exogenous focal point (highlighting one option), contrast it with the standard 4-category scale, and find that the coordination game approach is robust to competing exogenous sources of focality. Finally, Fallucchi et al. (2025) note that sample sizes for some of these studies are underpowered to detect medium-small effects, which could result in false positives.

Erkut et al. (2015) test whether ratings in the coordination game vary if provided by participants who read about the dictator game or just played it. They find that the coordination game produced consistent results regardless of the role assigned to the participants. Schmidt et al. (2020) find that there are differences in the reported norms elicited with the coordination game when lab subjects have prior experience with the dictator game but not when eliciting norms for everyday situations. Vesely and Klöckner (2018) and Chang et al. (2019) find that participants with different social identities (e.g., Republican versus Democrat) do not use subgroup-specific

---

<sup>5</sup>This could be mitigated by using a distribution measure that is corrected for the location within the interval. However, we are not aware of any studies that have done this.

shared perceptions in the coordination task unless their social identity is made salient to them in the experiment. [Vesely \(2015\)](#) and [Huffman et al. \(2024\)](#) use the coordination-game approach with and without incentives to elicit norms and find that monetary incentives did not change the qualitative results. [Fallucchi et al. \(2025\)](#) set out to use larger sample sizes to provide an analysis of the effects of incentivizing responses in the coordination game approach. They find modest improvements in response quality and show that incentives mitigate response biases that arise when participants’ self-serving motivations conflict with accurate responding. However, note that the vast majority of these sensitivity checks are conducted using the dictator game, which appears to have well-established strong norms, though [Fallucchi et al. \(2025\)](#) is an exception. [König-Kersting \(2024\)](#) concludes that “it has become clear that KW’s method is remarkably robust and insensitive to variations in task wording and the salience of incentives.”

### 3.3.2 Direct Comparisons of Sequential Opinion-Matching vs. Coordination Games

These two methods elicit different types of information about norms due to differences in the objectives they have for measurement. We discuss them briefly in turn. The sequential-opinion-matching approach aims to capture the beliefs necessary for a norm to exist (e.g., normative personal beliefs and expectations for action). This is why step 2 measures the predicted proportion of participants who believe each action is most appropriate. The coordination-game approach aims to capture the normative characteristics of each action (e.g., whether an action is appropriate or inappropriate for a decision situation). This is why the method measures the expected modal appropriateness of each action. Both use the data to construct an empirical proxy of the social norm. In the sequential opinion-matching approach, the measure is constructed using step-two responses, and in the coordination-game approach, it is constructed by taking the average rating for each action and building a “norm function” that maps appropriateness to each action rated.

[Lane et al. \(2023\)](#) compared (a variant of) the sequential opinion-matching and the coordination game and elicited consistent results across the two methods. To ease the comparison, they modified the sequential-opinion-matching response format to allow participants to rate each action on a four-point appropriateness scale (rather than the binary fair/not fair or should/should not scale). This could potentially account for the high degree of overlap between the methods, but more research is needed to understand the source of this overlap. [Aycinena et al. \(2024a\)](#) compared the two methods against a binarized scoring method (BSM) that elicited belief distributions while preserving incentive compatibility. [Aycinena et al. \(2024a\)](#) found that all three approaches identified externally validated norms, that the BSM approach did not produce the false-consensus effects that can threaten the sequential-opinion-matching approach, and that within a norm-dependent utility model, the coordination-game approach outperformed both BSM and sequential opinion-matching and outperformed BSM for predicting behavior.

Both approaches can be justified, depending on the research question, in contexts where there is a single norm and a generally shared belief about that norm. However, there are several reasons why such an assumption might not be justified. For example, people may not be sure about whether or which norm applies, a community may have sub-groups with different norms ([Groenendyk et al., 2023](#); [Panizza et al., 2024a](#)), or people believe that there is only one norm

but differ in their beliefs about what it actually is (Aycinena et al., 2022; Kimbrough et al., 2024). The first case can be thought of as “norm uncertainty” and is intuitively defined as a situation where people are not sure what the norm is. The latter examples are both cases where there are “multiple norms” and can intuitively be thought of as cases where there is disagreement about which norm applies. In the coordination-game approach developed by Krupka and Weber (2013), norm uncertainty could be inferred from the variance in ratings. To better interpret this variance, researchers can include follow-up questions, as in Krupka et al. (2017), that ask participants how sure they are of each rating guess. However, more recently, Schmidt (2025) extends the Krupka and Weber approach to capture norm uncertainty (we discuss this approach in section 3). In contrast, the sequential-opinion-matching approach developed by Bicchieri and Xiao (2009) uses a binary format, making it more difficult to capture uncertainty.

In the presence of multiple norms, using either of these methods becomes a little tricky to interpret. For example, if a coordination game results in two “peaks” (where two different actions are both considered appropriate), this could mean one of two things: either there is a single norm that allows for both actions equally, or there are actually two separate norms. In the second case, people who choose the first action may have systematically different views on the appropriateness of the other action compared to those who choose the second action. Furthermore, this method would not allow one to see the underlying normative beliefs out of which the distribution is composed (Kimbrough et al. (2024) and Panizza et al. (2024b) show evidence consistent with this latter point). To address these concerns, newer approaches focus on eliciting beliefs about distributions, which we review below. Finally, both approaches, for different reasons, would struggle to distinguish when personal normative beliefs diverge from group-level norms (Behnk et al., 2022), and neither would do well if, as Kimbrough et al. (2024) note, norms are in flux. We return to this point shortly.

## 4 New Experimental Techniques To Measure Social Norms

New approaches to norm measurement tended to focus on eliciting beliefs about distributions as this may be better suited to identify norm uncertainty or the presence of multiple norms.

The Binary Scoring Method (BSM), described in Aycinena et al. (2024a), is a two-step norm elicitation method. In Step 1, participants allocate 100 tokens across a set of actions according to their own beliefs about the “relative moral worthiness” of each action. In step 2, they predict how the tokens of 10 other randomly selected respondents are distributed, on average, across the actions. Payment to subjects is based on how closely their predicted distribution matches the actual distribution, with higher overlap resulting in a higher probability of payment (the Belief Elicitation by Superimposition Approach (BESA for short, Fragiadakis et al. (2019)).

Thus, the approach elicits a relative evaluation of each action because participants are constrained by the number of tokens they have. The approach does not give information on valence since the scale is constrained at zero tokens. However, one can empirically characterize the proportion who hold a particular belief and the distribution of those beliefs. The second step elicits beliefs about distributions. Because this approach relies on the first step of eliciting personal normative beliefs, it can be challenging to use when examining taboo topics, as subjects may



respond to demand effects.

The quadratic scoring rule (QSR) is another recent approach that focuses on eliciting the distribution of norm perceptions at the individual level but where responses can be aggregated to yield modal and mean normative beliefs (Dimant, 2023; Dimant et al., 2024b). In this task, one set of participants reads about a game and is asked what one should do and also plays the game. Another set of participants reads about the game and the set of actions that could be taken in that game. That set is then asked to indicate both their normative expectations (what they believe other participants said one should do in this game) and their empirical expectations (what they believe other participants have actually done in this game).

The novel feature of the approach is that the latter set of participants are asked to indicate their beliefs about the *distribution* of responses by participants who participated in the game. They do so by allocating a total of 10 points across the actions to indicate their beliefs about the distribution of normative beliefs (i.e., how other participants responded) and/or their beliefs about empirical expectations (i.e., how other participants behaved). The more likely they think that a previous participant said that one should take this action (or that a participant did take the action), the more points they should allocate to it. Figure 1 shows the generic graphical interface for this task for a public goods game where actions ranged from contributing 0 to 4 tokens. In addition, whether participants were indicating their normative or empirical distribution guesses, a second question asked for participants' confidence (0%–100%). Payoffs were calculated using a quadratic scoring rule. One randomly selected decision is payoff-relevant to reduce hedging:  $Q_j(p) = \alpha + 2\beta * p_j - \beta * \sum_{i=1}^n (p_i)^2$ , where  $p_j$  is the probability a participant assigns to the true option.<sup>6</sup>

0 token	1 token	2 token	3 token	4 token
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
+	+	+	+	+
-	-	-	-	-

**Figure 1:** The graphical interface of the QSR norm elicitation. Participants allocate a total of 10 points across actions ranging from contributing 0 to 4 tokens to a Public Goods Game. The more likely they think an action is, the more points they should allocate to it.

Thus, this approach elicits a relative evaluation because participants are constrained by the number of points they have to distribute across the actions. The approach does not give information on valence since the scale is constrained at zero points. The response format elicits individual beliefs about the distribution of beliefs (normative and empirical, depending on the question) and can be aggregated to give modes and means for those distributions.

QSR does not work well when one cannot objectively verify the true state. This makes it less suitable for measuring injunctive norms; rather, it measures beliefs about injunctive

<sup>6</sup>In the example,  $\alpha = \beta = 0.5$ , implying that participants earn between \$0 and \$1 depending on their stated beliefs and the truth.

norms (see the discussion above regarding methodological features of the Krupka-Weber task). However, it can be used well to elicit beliefs about others’ behaviors. Additionally, it is only incentive-compatible in risk-neutral conditions (Offerman et al., 2009; Charness et al., 2021). Notwithstanding, Dimant (2023) used data from Dimant et al. (2024b) and validated this method by demonstrating a strong relationship between the distribution of norm-related beliefs and observed behavior in the experiment.

Schmidt (2025) adapts the coordination game approach but modifies the underlying coordination game such that subjects are not forced to choose one appropriateness rating for each action. Rather, they can distribute a budget of tokens by placing those tokens on multiple appropriateness ratings at the same time. Specifically, subjects learn about mini-dictator games that have 2 actions (for example, one game yields a dictator/recipient payoff of 10,10 each). For each mini-dictator game, they rate how appropriate the allocation is using an ordinal rating scale (“very appropriate”, “somewhat appropriate”, “somewhat inappropriate” or “very inappropriate”). Each subject receives a budget of 100 tokens, and they must distribute all the tokens across the appropriateness ratings for that action. If the norm elicitation stage is selected for payment, then one of the mini-dictator games is drawn by chance, the experimenter calculates which alternative the subjects in that session assigned the most tokens to on average (the “winning alternative”), and each subject receives €0.20 per token assigned to that alternative. Thus, each subject earns between €0 (no tokens assigned to the winning alternative) and €20 (all 100 tokens assigned to the winning alternative).

Thus, the incentives are to guess where the most tokens will be placed, however, because respondents can spread tokens, it also can be used to identify uncertainty about the “correct” appropriateness rating for any action. The approach elicits a relative evaluation of each action because participants are constrained by the number of tokens they have. The approach gives information on valence since the tokens are allocated to appropriateness ratings. However, one can empirically characterize the proportion who hold a particular belief and the distribution of those beliefs. Except for the adapted choice set, the dynamics and incentives present in the coordination game approach are the same in both Schmidt’s and Krupka and Weber’s approach: subjects are rewarded for anticipating the modal behavior of others because they receive a profit that is proportional to the number of tokens that they assigned to the most popular alternative. However, unlike in Krupka and Weber, the dimensionality of the rating task and action space would increase quickly because subjects are no longer picking one appropriateness rating for an action, but are allocating tokens across the four ratings. The action space increases if one studies games that have more actions than the mini-dictator game studied here.

Norm multiplicity is another measurement challenge. Reuben and Riedl (2013) use survey data to show that uninvolved individuals hold well-defined yet conflicting normative views of fair contribution rules related to efficiency, equality, and equity in a public goods game. Kimbrough et al. (2024) employed the coordination game approach and then tested for the presence of multiple norms econometrically by analyzing individual-level variation in reported norms between people and within people across time. Using latent transition analyses, and extending latent class models to a panel setting, they showed that these variations have an interpretable structure that is consistent with the presence of multiple norms. Kimbrough et al. (2024) ad-

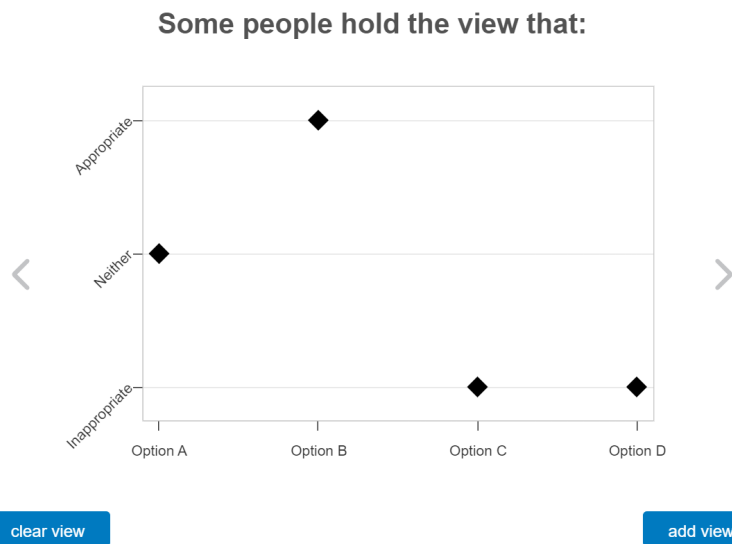


vised researchers to carefully conceptualize injunctive norms, even in well-studied contexts such as the dictator game, because the elicited belief profiles will have heterogeneity that can be meaningfully classified into distinct groups (Fromell et al., 2021).

To identify multiple norms directly, Panizza et al. (2024b) proposed a two-step Norm-Drawing Task (NDT). In step 1, participants learn about a situation and a set of actions. They are then asked to estimate how many distinct norms might be present in a given population and to create separate norm profiles for each, mapping actions to appropriateness ratings (e.g., “action 1 is very appropriate”, “action 2 is inappropriate”, “action 3 is neutral”). Figure 2 shows the interface participants used to create the profiles. They can create as many as they think might be present for that situation, but in the study by Panizza et al. (2024b) that we review, the appropriateness scale was restricted to three ratings, and participants were restricted to ten drawings. This process produces N distinct norm profiles. In step 2, participants review all the norm profiles they personally drew and indicate their guess about how many members out of 100 of an identified reference group (e.g., Democrats) would endorse that profile. The incentive scheme is to coordinate one’s drawings and the distribution of assigned points with other participants given the same task (the BESA approach, Fragiadakis et al., 2019). This is because the payment is determined by how well the resulting distribution compares to the average distribution guessed by other participants. Their probability of winning a prize is equal to the sum of the minimum between their guess and the average guess for each norm.

### 1. How many different views are there? What do these views look like?

You can make **between 1 and 10 drawings** to represent different views that you think exist.  
All actions need to have a rating.



**Figure 2:** The graphical interface for entering ratings for each action. Participants start with a blank drawing and can click or drag the dots to rate the appropriateness of each action. Participants can also navigate through the different drawings, delete drawings, or add up to ten different drawings. Actions are connected by lines if on a continuous scale; otherwise unconnected.

Thus, the response format asks subjects to draw profiles of norms; the step 1 task elicits an

absolute evaluation because each action in the drawing task can be evaluated independently. Nothing about the drawing tasks asks or incentivizes respondents to consider action ratings relative to each other. The responses yield positive and negative valence of the norm, and the second stage elicits beliefs about distributions. When studied at the individual level, the responses reflect participants' beliefs about others' normative beliefs and, when aggregated, the responses indicate how much agreement there is among participants about the number of normative beliefs. [Panizza et al. \(2024b\)](#) find that in both the dictator game and an allocation game, participants reported four to five normative views on average. The evidence of multiple norms in the dictator game may be surprising, but it supports the findings of [Kimbrough et al. \(2024\)](#) noted above and suggests that in more complex decision-making environments, one might expect and wish to measure the presence of multiple norms.<sup>7</sup> [Panizza et al. \(2024a\)](#) validate this method in the context of various contentious topics in the U.S., with a focus on the level of pluralism and (dis)agreement between Republicans and Democrats. The authors note that the NDT can also be used to detect pluralistic ignorance or polarization (see also, [Iyengar et al., 2019](#); [Bursztyn et al., 2020b](#); [Levy, 2021](#); [Dimant, 2024](#)).

The NDT, therefore, has advantages that can be valuable if the researcher is interested in measuring norm perceptions without defining what groups exist in the population. The latter is useful when it is unclear which members of the population share the same views. It also can be applied in cases where there are clearly identifiable groups, but the researcher believes that there exists within-group norm heterogeneity or that known groups are composed of unknown subgroups. Furthermore, it can be used to identify polarization or pluralistic ignorance.

However, as the NDT is a generalization of the Krupka-Weber task, it has the (same) feature of multiple equilibria and is predicated on the same intuition that shared awareness of the set of norms in the population and their relative representation can help resolve the coordination problem. However, some experimental design choices enhance coordination but may also limit the ability to capture the complexity they aim to reveal. For example, in [Panizza et al. \(2024b\)](#), participants rate each action's appropriateness on a 3-item Likert scale (appropriate, neither appropriate nor inappropriate, inappropriate) and, the authors note in a footnote that they ran a series of calibration pilots which suggest that a higher number of options would reduce participants' ability to coordinate in the specific scenarios tested. In addition, the NDT may not be well suited for situations where the number of possible norms is large. The task asks participants to draw what they think are the possible norms and respondents may run into fatigue as they work their way through all the possible norms they think may apply. In addition, in the coordination-game method, the number of games is always equal to the number of actions over which games are played, whereas in the NDT the number of games in the second stage is endogenous because it depends on how many profiles were drawn in Step 1.<sup>8</sup>

---

<sup>7</sup>Though the presence of multiple norms was not identified in earlier papers, there was some evidence that this might be occurring. For example, there was a large fraction of subjects who chose the equal split in [Krupka and Weber \(2013\)](#) and a larger group that chose to keep all of the endowment.

<sup>8</sup>This can be addressed by limiting the number of norms one can draw as [Panizza et al. \(2024a\)](#) do in their study.

## 4.1 Complexity of Implementation

So far, we have not given much attention to implementation complexity. “Ease of Implementation” is perhaps the most subjective element on which to characterize these methods. We will limit our evaluation of implementation, from less complex to more complex, based on three features: an experimental protocol can easily be converted across formats (e.g., paper-pencil to online), the ease with which instructions, also known as ‘the task’, are explained to participants, and ease of understanding incentives. While there may be other features of any particular protocol reviewed here that would reduce or increase complexity, these dimensions are ones that we imagine experimenters would wish to consider and, as such, will be our evaluative lens.

**Least Complex Elicitation Method:** For reference, we consider surveys as our benchmark. These are not strictly used in the studies we review here, but the methodology offers a good starting point for our assessment of ease: instructions explaining how to respond are typically short/easy to understand, respondents have familiarity with the response format, and questions are typically intuitively phrased and understood. It can easily be adapted to either paper-pencil or computerized formats, and the process for data collection is not onerous for the researcher. Typically, incentives are paid for participation or completion.

**Less Complex Elicitation Method:** While several approaches might meet our first two criteria (conversion to different formats and ease of understanding the task), all of the current approaches rely on incentive structures that are conditional on responses. Most also use incentive structures that entail explaining complex conditional payment structures such as matching on modal ratings or using quadratic scoring rules. Thus, we deem no current approach as less complex when assessed on these dimensions.

**Somewhat Complex Elicitation Method:** [Bicchieri and Xiao \(2009\)](#), [Aycinena et al. \(2024a\)](#), [Dimant \(2023\)](#), and [Schmidt \(2025\)](#) might all be considered somewhat complex but for different reasons. In [Bicchieri and Xiao \(2009\)](#) and [Aycinena et al. \(2024a\)](#) the first step is arguably rather intuitive (asking subjects whether they think dictators should split the money). However, in order to pay subjects based on their responses, the second step involves some form of aggregating responses from the first step and this typically means the procedure cannot be done easily in a paper and pencil format. The approach also relies on more sophistication on the part of subjects to understand what is asked of them on the second task where they are paid if they accurately guess what most others do or disapprove of doing. In [Aycinena et al. \(2024a\)](#), the payment might be described as a bit more complex because it is based on how closely participants’ predicted distribution matches the actual distribution of responses, with higher overlap resulting in a higher probability of payment. [Dimant \(2023\)](#) and [Schmidt \(2025\)](#) introduce methods that elicit distributions of norm-related beliefs. These approaches, though not yet studied using paper and pencil, would likely not be difficult to convert. The tasks ask participants to respond in a way that maps to commonly used notions - allocating more points indicates more people believe or do a certain thing. However, the payment mechanisms require some sophistication on the part of the participant. The quadratic scoring rule (QSR) is complex, and [Schmidt et al. \(2020\)](#) notes that the instructions were detailed and provided many examples.

**More Complex Elicitation Method:** [Krupka and Weber \(2013\)](#) and [Panizza et al.](#)

(2024b) are among the more complex methods we review. In [Krupka and Weber \(2013\)](#), although the process consists of a single step where participants read about a situation, the task itself may not be intuitive. Participants need to coordinate on appropriateness ratings, which use an ordinal scale that requires additional explanation (e.g., clarifying that “socially appropriate” means what most others think one ought to do). In addition, the payment structure, which rewards modal responses, demands a level of sophistication from participants. While the method elicits ratings in one step, researchers must later identify the modal responses to determine payments (or draw a random response), making it difficult to implement in a paper-and-pencil format. Although the original [Krupka and Weber \(2013\)](#) study used paper and pencil, payments were delayed because responses had to be aggregated manually before final compensation. [Panizza et al. \(2024b\)](#) introduced a two-step norm-drawing task (NDT), which is potentially easier to adapt to a paper-and-pencil format. However, participants are asked to generate as many norm profiles as they find necessary, which could create practical challenges. For example, prolific participants flipping through pages could distract others, and providing multiple sheets for drawings might introduce experimenter demand effects. Step 1 also requires participants to grasp a somewhat complex task. Finally, the incentive structure adds another layer of sophistication, as participants’ chances of winning a prize depend on how closely their guesses match the average distribution of norm profiles drawn by others.

## 5 Discussion and Conclusion

Table 1 summarizes the discussion above into a few important elements of each approach that the researcher ought to consider when deciding between these methods. The table identifies the best paper to learn about the task in column 1, provides a subjective assessment of implementation complexity, provides a characterization of the responses obtained with each method, an overview of some of the robustness tests, and lists a strength and a weakness for each method (though many more are covered in the text above).

**Table 1:** Comparison of Norm Elicitation Tasks

<i>Norm Elicitation Task</i>	<i>Implementation Complexity</i>	<i>Response Characteristics</i>	<i>Robustness &amp; Scope</i>	<i>One Strength &amp; One Limitation</i>
<i>Bicchieri and Xiao (2009)</i>	Somewhat	Rate on binary scale, absolute rating on action, pos. & neg. norm valence, step 2 elicits central tendency beliefs	Not yet systematically validated, frequently used both with & without monetary incentives. Some evidence of vulnerability to false consensus, Step 1 beliefs susceptible to social-desirability bias	Can discover norms and measure pluralistic ignorance. Challenging to use for taboo topics due to demand effect; when all actions are acceptable may yield false positives
<i>Krupka and Weber (2013)</i>	More	Rate actions on ordinal scale, absolute rating on each action, pos. & neg. norm valence, elicits beliefs about central tendency	Robust results with & without incentives, to between- & within-subject elicitation, to competing sources of focality. Reference groups matter. Low rate of false-positives	Integrates with models of norm-dependent utility, can estimate the weight a norm has in decision-making. Multiple norms are hard to discover
<i>Dimant (2023)</i>	Somewhat	Rate actions using a fixed budget, relative rating on each action, no valence, elicits beliefs about distributions	Elicits individual-level beliefs about distribution of normative and/or empirical expectations, aggregated yields modal & mean beliefs, can be insightful studying polarization & norms in strategic settings	Useful when distribution of norm-related beliefs (e.g., uncertain environments) are desired. Norm discovery is limited and may be distortions absent risk neutrality
<i>*Aycinena et al. (2024a)</i>	Somewhat	Rate actions using a fixed budget, relative rating on each action, no valence, elicits beliefs about distributions	Not yet evaluated for robustness	Enables analysis of the relative intensity/ strength of normative beliefs. Challenging to use when examining taboo topics due to demand effects
<i>*Panizza et al. (2024b)</i>	More	Draw profiles of norms, absolute rating on actions, pos. & neg. norm valence, elicits beliefs about distributions	Some evidence with & without incentives. Elicits multiple norms	Endogenously identify how many norms believed to exist. Without restrictions on the drawing task and appropriateness categories, dimensionality is a challenge
<i>Schmidt (2025)</i>	Somewhat	Rate actions using a fixed budget, relative rating on action, pos. & neg. norm valence, elicits beliefs about distributions	Not yet evaluated for robustness	Can be used to identify norm uncertainty. The dimensionality of the rating task and action space increases quickly

**Note:** This table summarizes a few of the highlights from the discussion of each method above. The \* indicates that the paper has not yet been peer-reviewed. Ready-made scripts for the norm elicitation methods discussed in this paper are available for download via our OSF repository: [link](#).

One theme that emerges from this review is that early methods focused on eliciting mean or modal normative beliefs. However, recent work has shifted its focus to eliciting beliefs about the distribution of normative beliefs which allows the researcher to draw inferences on the degree of uncertainty that underlies norm assessments. This shift in research focus has been accompanied by some exciting developments that have innovated on the early methodological “parents”.

Aycinena et al. (2024a) and Dimant (2023) share roots with Bicchieri and Xiao (2009) in that they first elicit subjects’ personal normative beliefs about the fairness or appropriateness of actions. Aycinena et al. (2024a) and Dimant (2023) innovate on the method by then eliciting distributions of beliefs but in the former, payment is determined using BESA, while in the latter, it is determined by the quadratic scoring rule. On the other hand, Panizza et al. (2024b) and Schmidt et al. (2022) share roots with Krupka and Weber (2013) but Panizza et al. (2024b) uses the BESA approach to generalize the coordination game method from second-order to third-order beliefs and Schmidt (2025) essentially has a similar incentive structure to Krupka and Weber (2013). However, while both Schmidt (2025) and Dimant (2023) elicit distributions, subjects are trying to bet on a unique outcome. The more points subjects put on the “correct” point (which is unknown to the individual at the time of making their guess by stating a distribution), the higher the payoff. As Schmidt (2025) so clearly explains, “...in both cases....the correct option corresponds to one single alternative, and subjects can bet on that option while having the possibility to diversify. As a side effect, the experimenter learns about the underlying degree of uncertainty...”. Contrast this with the task set to subjects in Aycinena et al. (2024a) and Panizza et al. (2024b), who are incentivized to guess the actual correct *distribution*.

All of these are important and exciting advances because they allow us to better discover cases where there may be uncertainty or where there are multiple (competing) norms. Very little work has focused on norm change or norm emergence, and, together with the early methods, these will provide new ways for researchers to understand how and when norms are created, changed, passed on, or destroyed. Finally, it is worth pointing out that these approaches offer techniques that can be combined to develop new methods for norm measurement that have their own advantages and disadvantages.

## References

- Abeler, J., Nosenzo, D., and Raymond, C. (2019). Preferences for truth-telling. *Econometrica*.
- Akerlof, G. A. (1980). A theory of social custom, of which unemployment may be one consequence. *The Quarterly Journal of Economics*, 94(4):749–775.
- Akerlof, G. A. and Kranton, R. E. (2000). Economics and identity. *The Quarterly Journal of Economics*, 115(3):715–753.
- Alesina, A. and Angeletos, G.-M. (2005). Fairness and redistribution. *American Economic Review*, 95(4):960–980.
- Almås, I., Cappelen, A. W., and Tungodden, B. (2020). Cutthroat capitalism versus cuddly socialism: Are americans more meritocratic and efficiency-seeking than scandinavians? *Journal of Political Economy*, 128(5):1753–1788.
- Andreoni, J. and Bernheim, B. D. (2009). Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects. *Econometrica*, 77(5):1607–1636.
- Axelrod, R. (1986). An evolutionary approach to norms. *American Political Science Review*, 80(4):1095–1111.
- Aycinena, D., Bogliacino, F., and Kimbrough, E. O. (2024a). Measuring norms: A comparison of the predictive and descriptive power of three methods. *Working Paper*.
- Aycinena, D., Bogliacino, F., and Kimbrough, E. O. (2024b). Measuring norms: Assessing the threat of social desirability bias to the bicchieri and xiao elicitation method. *Journal of Economic Behavior & Organization*, 222:225–239.
- Aycinena, D., Bogliacino, F., and Kimbrough, E. O. (2024c). Measuring norms: Eliciting normative expectations with krupka and weber’s method allowing for neutral reports. *Working Paper*.
- Aycinena, D., Rentschler, L., Beranek, B., and Schulz, J. F. (2022). Social norms and dishonesty across societies. *Proceedings of the National Academy of Sciences*, 119(31):e2120138119.
- Bae, J., Choi, B., Krupka, E., and Lee, S. (2025). Harnessing project identity and safety norms to promote construction workers’ safety behavior: A field intervention study. *Journal of Construction Engineering and Management*.
- Bašić, Z. and Verrina, E. (2024). Personal norms—and not only social norms—shape economic behavior. *Journal of Public Economics*, 239:105255.
- Behnk, S., Hao, L., and Reuben, E. (2022). Shifting normative beliefs: On why groups behave more antisocially than individuals. *European Economic Review*, 145:104116.
- Benabou, R. and Tirole, J. (2006). Belief in a just world and redistributive politics. *The Quarterly Journal of Economics*, 121(2):699–746.

- Bicchieri, C. (2006). *The grammar of society: The nature and dynamics of social norms*. Cambridge University Press.
- Bicchieri, C. and Chavez, A. (2010). Behaving as expected: Public information and fairness norms. *Journal of Behavioral Decision Making*, 23(2):161–178.
- Bicchieri, C. and Dimant, E. (2019). Nudging with care: The risks and benefits of social information. *Public choice*, 191(3):443–464.
- Bicchieri, C., Dimant, E., Gächter, S., and Nosenzo, D. (2022). Social proximity and the erosion of norm compliance. *Games and Economic Behavior*, 132:59–72.
- Bicchieri, C., Dimant, E., Gelfand, M., and Sonderegger, S. (2023). Social norms and behavior change: The interdisciplinary research frontier.
- Bicchieri, C., Dimant, E., and 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., and Sontuoso, A. (2018). Social norms. *The Stanford encyclopedia of philosophy*.
- Bicchieri, C. and Sontuoso, A. (2020). Game-theoretic accounts of social norms: The role of normative expectations. In *Handbook of experimental game theory*, pages 241–255. Edward Elgar Publishing.
- Bicchieri, C. and Xiao, E. (2009). Do the right thing: but only if others do so. *Journal of Behavioral Decision Making*, 22(2):191–208.
- Binmore, K. and Samuelson, L. (2006). The evolution of focal points. *Games and Economic Behavior*, 55(1):21–42.
- Burks, S. V. and Krupka, E. L. (2012). A multimethod approach to identifying norms and normative expectations within a corporate hierarchy: Evidence from the financial services industry. *Management Science*, 58(1):203–217.
- Bursztyn, L., Egorov, G., and Fiorin, S. (2020a). From extreme to mainstream: The erosion of social norms. *American Economic Review*, 110(11):3522–48.
- Bursztyn, L., González, A. L., and Yanagizawa-Drott, D. (2020b). Misperceived social norms: Women working outside the home in saudi arabia. *American Economic Review*, 110(10):2997–3029.
- Carpenter, J. and Robbett, A. (2024). Measuring socially appropriate social preferences. *Games and Economic Behavior*, 147:517–532.
- Chang, D., Chen, R., and Krupka, E. (2019). Rhetoric matters: A social norms explanation for the anomaly of framing. *Games and Economic Behavior*, 116:158–178.
- Charness, G., Gneezy, U., and Rasocha, V. (2021). Experimental methods: Eliciting beliefs. *Journal of Economic Behavior & Organization*, 189:234–256.



- Choi, B., Ahn, S., and Lee, S. (2017). Construction workers’ group norms and personal standards regarding safety behavior: Social identity theory perspective. *Journal of management in engineering*, 33(4):04017001.
- Cialdini, R. B., Reno, R. R., and Kallgren, C. A. (1990). A focus theory of normative conduct: recycling the concept of norms to reduce littering in public places. *Journal of personality and social psychology*, 58(6):1015.
- Coleman, J. S. (1994). *Foundations of social theory*. Harvard university press.
- Dimant, E. (2023). Beyond average: A method for measuring the tightness, looseness, and polarization of social norms. *Economics Letters*, 233:111417.
- Dimant, E. (2024). Hate trumps love: The impact of political polarization on social preferences. *Management Science*, 70(1):1–31.
- Dimant, E., Galeotti, F., and Villeval, M. C. (2024a). Motivated information acquisition and social norm formation. *European Economic Review*.
- Dimant, E., Gelfand, M., Hochleitner, A., and Sonderegger, S. (2024b). Strategic behavior with tight, loose, and polarized norms. *Management Science*.
- Dimant, E. and Gesche, T. (2023). Nudging enforcers: How norm perceptions and motives for lying shape sanctions. *PNAS Nexus*, 2(7):pgad224.
- Dimant, E., Reinhardt, L., and Sambanis, N. (2024c). Racial identity, reactions to inequality, and fairness concepts among Americans. Working paper.
- Elster, J. (1989). Social norms and economic theory. *The Journal of Economic Perspectives (1986-1998)*, 3(4):99.
- Elster, J. (1991). Rationality and social norms. *European Journal of Sociology/Archives Europeennes de Sociologie*, 32(1):109–129.
- Erkut, H., Nosenzo, D., and Sefton, M. (2015). Identifying social norms using coordination games: Spectators vs. stakeholders. *Economics Letters*, 130:28–31.
- Fallucchi, F., Fromell, H., and Nosenzo, D. (2025). Incentivizing social norm elicitation. Working paper.
- Fallucchi, F. and Nosenzo, D. (2022). The coordinating power of social norms. *Experimental Economics*, 25(1):1–25.
- Fehr, E. and Charness, G. (2025). Social preferences: Fundamental characteristics and economic consequences. *Journal of Economic Literature*.
- Fehr, E. and Gächter, S. (2000a). Cooperation and punishment in public goods experiments. *American Economic Review*, 90(4):980–994.

- Fehr, E. and Gächter, S. (2000b). Fairness and retaliation: The economics of reciprocity. *Journal of economic perspectives*, 14(3):159–181.
- Fehr, E. and Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The quarterly journal of economics*, 114(3):817–868.
- Fehr, E. and Schurtenberger, I. (2018). Normative foundations of human cooperation. *Nature Human Behaviour*, 2(7):458.
- Fershtman, C., Gneezy, U., and Hoffman, M. (2011). Taboos and identity: Considering the unthinkable. *American Economic Journal: Microeconomics*, 3(2):139–164.
- Fragiadakis, D. E., Kovaliukaite, A., and Rojo Arjona, D. (2019). the belief elicitation by superimposition approach. Working paper.
- Fromell, H., Nosenzo, D., Owens, T., and Tufano, F. (2021). One size does not fit all: Plurality of social norms and saving behavior in Kenya. *Journal of Economic Behavior & Organization*, 192:73–91.
- Gelfand, M. J., Gavrillets, S., and Nunn, N. (2024). Norm dynamics: Interdisciplinary perspectives on social norm emergence, persistence, and change. *Annual Review of Psychology*, 75.
- Gneezy, U. (2005). Deception: The role of consequences. *American Economic Review*, 95(1):384–394.
- Groenendyk, E., Kimbrough, E. O., and Pickup, M. (2023). How norms shape the nature of belief systems in mass publics. *American Journal of Political Science*, 67(3):623–638.
- Güth, W., Schmittberger, R., and Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of economic behavior & organization*, 3(4):367–388.
- Hardin, R. (1982). *Collective Action*. Resources for the Future.
- Heinicke, F., König-Kersting, C., and Schmidt, R. (2022). Injunctive vs. descriptive social norms and reference group dependence. *Journal of Economic Behavior & Organization*, 195:199–218.
- Horne, C. and Mollborn, S. (2020). Norms: An integrated framework. *Annual Review of Sociology*, 46(1):467–487.
- Huffman, D., Kohno, G., Madies, P., Vogrinec, S., Wang, S. W., and Yagnaraman, D. (2024). Measuring social norm variation across contexts: Replication and comparison to alternative methods.
- Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., and Westwood, S. J. (2019). The origins and consequences of affective polarization in the United States. *Annual Review of Political Science*, 22:129–146.
- Kahneman, D., Knetsch, J. L., Thaler, R., et al. (1986). Fairness as a constraint on profit seeking: Entitlements in the market. *American economic review*, 76(4):728–741.

- Kimbrough, E. O., Krupka, E. L., Kumar, R., Murray, J. M., Ramalingam, A., Sánchez-Franco, S., Sarmiento, O. L., Kee, F., and Hunter, R. F. (2024). On the stability of norms and norm-following propensity: A cross-cultural panel study with adolescents. *Experimental Economics*, pages 1–28.
- Kimbrough, E. O. and Vostroknutov, A. (2016). Norms make preferences social. *Journal of the European Economic Association*, 14(3):608–638.
- König-Kersting, C. (2024). On the robustness of social norm elicitation. *Journal of the Economic Science Association*, pages 1–13.
- Krupka, E. and Weber, R. A. (2009). The focusing and informational effects of norms on pro-social behavior. *Journal of Economic psychology*, 30(3):307–320.
- Krupka, E. L., Leider, S., and Jiang, M. (2017). A meeting of the minds: informal agreements and social norms. *Management Science*, 63(6):1708–1729.
- Krupka, E. L. and 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.
- Lane, T., Nosenzo, D., and Sonderegger, S. (2023). Law and norms: Empirical evidence. *American Economic Review*, 113(5):1255–1293.
- Levy, R. (2021). Social media, news consumption, and polarization: evidence from a field experiment. *American Economic Review*, 111(3):831–70.
- Mehta, J., Starmer, C., and Sugden, R. (1994). The nature of salience: An experimental investigation of pure coordination games. *The American Economic Review*, 84(3):658–673.
- Merguei, N., Strobel, M., and Vostroknutov, A. (2022). Moral opportunism as a consequence of decision making under uncertainty. *Journal of Economic Behavior & Organization*, 197:624–642.
- Nosenzo, D. and Görges, L. (2020). Measuring social norms in economics: Why it is important and how it is done. *Analyse & Kritik*, 42(2):285–312.
- Offerman, T., Sonnemans, J., Van de Kuilen, G., and Wakker, P. P. (2009). A truth serum for non-bayesians: Correcting proper scoring rules for risk attitudes. *The Review of Economic Studies*, 76(4):1461–1489.
- Ostrom, E. (2000). Collective action and the evolution of social norms. *Journal of economic perspectives*, 14(3):137–158.
- Panizza, F., Dimant, E., Kimbrough, E. O., and Vostroknutov, A. (2024a). Measuring norm pluralism and perceived polarization in us politics. *PNAS nexus*, 3(10):pgae413.
- Panizza, F., Dimant, E., Kimbrough, E. O., and Vostroknutov, A. (2024b). Pluralism breeds tolerance. *Working Paper*. Available at SSRN: <https://dx.doi.org/10.2139/ssrn.4649792>.

- Reuben, E. and Riedl, A. (2013). Enforcement of contribution norms in public good games with heterogeneous populations. *Games and Economic Behavior*, 77(1):122–137.
- Schelling, T. C. (1980). *The Strategy of Conflict: with a new Preface by the Author*. Harvard university press.
- Schmidt, R. (2025). Social norm uncertainty: Measurement using coordination games and behavioral relevance. *Journal of Economic Behavior & Organization*, 232:106937.
- Schmidt, R., Heinicke, F., and König-Kersting, C. (2022). Using coordination games to measure beliefs. *Economics Letters*, 219:110821.
- Schmidt, R., Schwieren, C., and Sproten, A. N. (2020). Norms in the lab: Inexperienced versus experienced participants. *Journal of Economic Behavior & Organization*, 173:239–255.
- Schmidt, R. J. (2019). Identifying the ranking of focal points in coordination games on the individual level.
- Schwartz, S. H. and Fleishman, J. A. (1982). Effects of negative personal norms on helping behavior. *Personality and Social Psychology Bulletin*, 8(1):81–86.
- Sliwka, D. (2007). Trust as a Signal of a Social Norm and the Hidden Costs of Incentive Schemes. *American Economic Review*, 97(3):999–1012.
- Sugden, R. (1998). Normative expectations: the simultaneous evolution of institutions and norms. *Economics, values, and organization*, 73.
- Vesely, Š. (2015). Elicitation of normative and fairness judgments: Do incentives matter? *Judgment and Decision making*, 10(2):191–197.
- Vesely, S. and Klöckner, C. A. (2018). Global social norms and environmental behavior. *Environment and Behavior*, 50(3):247–272.
- Young, H. P. (2015). The evolution of social norms. *Annu. Rev. Econ*, 7:359–87.