THE RATIONALIZABILITY OF SURVEY RESPONSES

JOSE APESTEGUIA[†] AND MIGUEL A. BALLESTER[‡]

Abstract. We propose and study the concept of survey rationalizability. In the simplest scenario of dichotomous attitudinal surveys, survey rationalizability means that both questions and individuals' views can be positioned on the real line in such a way that individuals endorse only the questions that closely align with their views. We demonstrate how the relative positioning of questions can be learned through a revelation mechanism involving pairs of individuals and triplets of questions. We also establish that the acyclicity of these revelations is necessary and sufficient for rationalizability. Additionally, we show that our analysis readily extends to polytomous surveys and probabilistic data. Furthermore, we investigate the identification of the parameters in these models and prove that even the cardinal locations can be fully determined in an exponential version of the probabilistic model. Finally, we conclude by examining an alternative model of survey responses for aptitudes.

Keywords: Surveys; Rationalizability; Attitudes; Aptitudes.

JEL classification numbers: C02; C83; D01.

1. Introduction

Surveys are becoming increasingly popular in economics as tools for gathering data on various economic variables, such as subjective expectations, happiness, contingent valuations, beliefs, and political attitudes. Understanding the rationality foundations of survey responses is crucial for designing more accurate and informative surveys,

Date: May, 2023.

[*] We thank Marc Fleurbaey, Junnan He, Philippe Jehiel, Jean-François Laslier, Gaël Le Mens and Antonin Macé for helpful comments. Financial support by FEDER/Ministerio de Ciencia e Innovación (Agencia Estatal de Investigación) through Grant PID2021-125538NB-I00 and through the Severo Ochoa Programme for Centers of Excellence in R&D (Barcelona School of Economics CEX2019-000915-S), and Balliol College is gratefully acknowledged. Miguel Ballester is grateful to the Paris School of Economics for their hospitality during his visit.

[†]ICREA, Universitat Pompeu Fabra and BSE. E-mail: jose.apesteguia@upf.edu.

[‡]University of Oxford. E-mail: miguel.ballester@economics.ox.ac.uk.

ultimately leading to better decision-making and policies. This paper takes a first step in this direction by showing that the empirical content of classical survey response theories in the psychology literature can be studied using concepts and techniques inspired by traditional revealed preference analysis.

There are two main types of surveys: those that measure attitudes and those that measure aptitudes. We begin with attitudinal surveys, which present more challenges from a rationality perspective. The underlying assumption in much of the literature is that responses reveal the attitudes of individuals on a common spectrum. In psychology, a classical treatment of this concept is the item response theory with ideal points.[1] According to this approach, both survey questions and respondents' ideal points are positioned on a common scale, and individuals express greater support for questions closer to their ideal points. We commence our analysis by studying dichotomous surveys, which consist of yes/no questions. In this case, the above concept of endorsement-by-proximity can be modeled using the upper contour set of a strictly quasi-concave utility, which takes the form of an interval centered around the individual's ideal point. We say that survey responses are rationalizable if every question can be placed on the real line and each individual can be represented by a strictly quasi-concave utility and a threshold level such that questions with utility above the threshold level, and only those questions, are endorsed.

We show that the rationalizability of a dichotomous survey can be built on the basis of an intuitive revelation principle involving triplets of questions and pairs of individuals. We start by expanding the informational content of the survey by incorporating the hypothetical responses of fictitious individuals, in the following way. Consider two individuals with overlapping sets of responses. Given the model, their responses must correspond to two overlapping intervals in the real line. Since the union and difference of overlapping intervals of the real line are also intervals, we can understand the corresponding union and differences of their responses as the responses of some fictitious individuals. We then derive our revelation principle from the expanded survey. Suppose that we observe two individuals $\{n_a, n_b\}$ and three questions $\{q_a, q_I, q_b\}$ such that individual n_a endorses $\{q_a, q_I\}$ while individual n_b endorses $\{q_I, q_b\}$. For these responses to be consistent with the notion of endorsement-by-proximity, it must be that the question endorsed by both individuals, q_I, occupies the intermediate position

[1] See Thurstone (1928) and especially the unfolding theory of Coombs (1964) for early developments, and see Tay and Ng (2018) and Van der Linden (2018) for more recent accounts.

within the triplet. It is in this sense that we say that q_I is revealed to be intermediate in $\{q_a, q_I, q_b\}$. Theorem 1 shows that the acyclicity of the intermediateness revelations contained in the expanded survey is a necessary and sufficient condition for rationalizability.

We then show that our techniques in the analysis of dichotomous surveys prove useful in the study of more general settings. We consider two such settings; polytomous surveys, in which questions can be answered using one out of multiple levels of endorsement, and probabilistic survey data, where we allow for noisy responses. After adjusting our notion of survey rationalizability to consider monotone collections of thresholds, one for each endorsement level, and random thresholds, respectively, Theorems 2 and 3 establish the corresponding characterizations. Both results follow by simply transforming the richer survey datasets involved in these two settings into analytically equivalent dichotomous datasets, allowing us to recur to Theorem 1.

We then shift our focus to studying the identification of the models. Since they are ordinal, the key step in the analysis is to determine the order of questions and ideal points. When the survey data is rationalizable, the intermediateness revelations can be used to learn the order of questions, and in Proposition 1 we show that a richness condition guarantees that this order is unique. Regarding individuals' ideal points, Proposition 2 demonstrates that it is possible to establish bounds for them, with precise information being conveyed by probabilistic surveys, whereas polytomous surveys provide more information than dichotomous surveys. Furthermore, we demonstrate that specific probabilistic models can yield cardinal information. We postulate a simple parametric version of the model in which the probability of endorsing a question is an exponential function of its distance from the individual's ideal point and show, in Proposition 3, that the location of all questions and individuals' ideal points can be cardinally and uniquely identified.

Finally, we briefly analyze surveys oriented to the study of cumulative aptitudes, such as ability or health status, rather than attitudes. A common approach in the psychology literature is to apply dominance item response theory.[2] The dichotomous version of this approach is commonly known as Guttman scale. According to this, the difficulty of questions and the aptitude levels of individuals can be represented on the real line, with individuals responding positively to questions that are below their

[2]Classical treatments are Guttman (1944) and Rasch (1961); see Van der Linden (2018) for a more recent discussion.

aptitude levels. The study of the rationalizability in a Guttman scale is facilitated by its cumulative nature. Rather than working with the ternary intermediateness relation, we can simply work with the binary relation *more complex than* over the set of questions, allowing us to obtain a characterization result based upon the acyclicity of this binary revelation.

2. Related Literature

The empirical literature using surveys in economics is large and growing. Here, our focus is on theoretical and methodological treatments of survey responses in economics. Stantcheva (2022) gives a complete guide on the issues encountered when actually running a survey study, including the design of questions and the analysis of responses. Bertrand and Mullainathan (2001) proposes an econometric-based framework that accounts for errors in responses, enabling a meaningful interpretation of responses that are subject to cognitive biases. Falk, Neuber and Strack (2021) develops an individual-response model based on imperfect self-knowledge, where individuals' responses depend on a combination of private signals and the population mean. Benjamin, Guzman, Fleurbaey, Heffetz and Kimball (2023) proposes a methodology to uncover the informational content of self-reported well-being surveys, considering various potential response biases. Our contribution to this literature is to offer a theoretical framework that studies the rationality of survey responses founded in classical accounts in the psychology of survey responses. In addition, our probabilistic models of survey responses enable systematic consideration of various types of errors.

The endorsement of questions is related to various notions of approval. In a recent paper, Manzini, Mariotti and Ülkü (2022) consider the idea of "wishlisting," where an individual expresses interest or positive attitudes toward a subset of available items. They study the psychological foundations of an individual model of sequential approval. Another instance is approval voting. Interestingly, Laslier (2009) and Alós-Ferrer and Buckenmaier (2019) have shown that the equilibrium behavior in approval voting games may have the upper-contour-set structure analyzed in this paper.[3] Thus,

[3]Núñez and Xefteris (2017) study approval voting under single-peaked preferences, which is related to our notion of endorsement-by-proximity. They show that every Nash-implementable welfare optimum can indeed be implemented by means of approval voting mechanisms.

our main axiom, SARI, may be useful in providing a necessary condition for the study of equilibrium behavior under approval voting.

There is a long tradition in the political sciences of elucidating the ideological spectrum of politicians, voters, and policies (see, e.g., the textbook treatment of Poole and Rosenthal, 2007). In a recent paper, Barberá (2014) estimates the ideology of Twitter profiles by using a parametric model of item response theory with ideal points and connectivity data. Linking with another Twitter profile can be considered analogous to endorsing a survey question. We contribute to this literature by primarily studying nonparametric response models and offering testability conditions.

In many economic studies, online reviews are modeled as ordered choices in which a higher utility leads to a higher rating (see Greene and Hensher (2010) for a textbook discussion). Recently, Acemoglu, Makhdoumi, Malekian, and Ozdaglar (2022) proposed a social learning model that uses a response model based on thresholds to study the conditions under which the product quality can be learned. Our paper contributes to this literature by determining the conditions under which survey responses can be rationalized using a response model based on thresholds. We differ in our focus, as we investigate a unified scale where all attitudes can be positioned, rather than the transmission of information through responses.

Our techniques in the study of the rationalizability of dichotomous surveys are related to various literatures across scientific disciplines, such as the use of sortability and seriation in archaeology, anthropology, biology, computer science or psychology. The connection with all these fields is the so-called consecutive-ones-property, or simply c1p, of a $0-1$ matrix. C1p requires the existence of a permutation of the columns of the matrix such that the ones of every row become consecutive.[4] Our Theorem 1 contributes to this literature by providing a characterization of c1p based on the acyclicity of a ternary relation derived from matrix information. We achieve this objective by first expanding the information contained in the matrix and then constructing the ternary relation based on this expanded information.

[4]As an illustration of an application of c1p in a different field, consider the following statistical archaeology problem (Kendall (1969)). Data is given by a set of graves and a set of objects that are or are not in these graves. C1p is used to determine whether graves and objects can be located in the time-scale such that each grave contains only objects that were created, but were not obsolete, in the relevant period. See Hubert (1974) for a connection between these techniques and survey responses, and Liiv (2010) for an overview of their use across different disciplines.

Finally, we build our revealed analysis on the basis of an intermediateness ternary relation constructed from the responses of the individuals. Huntington and Kline (1917) and Fishburn (1971) represent early treatments of abstract ternary relations, with the purpose of representing the intermediateness of the real line. Our results contribute to this literature in a conceptual and technical way. As for the former, we apply the notion of intermediateness to a concrete economic problem deriving the ternary relation from data, and as for the latter, we show that one can extend certain incomplete ternary relations to linear ternary relations by way of using an acyclity condition.[5]

3. Rationalizability of attitudinal surveys: the dichotomous case

Let $\mathcal{Q} = \{1, \ldots, q, \ldots, Q\}$ be a set of questions on which a set of individuals $\mathcal{N} = \{1, \ldots, n, \ldots, N\}$ is surveyed. We start analyzing the case in which questions are dichotomous. Accordingly, a dichotomous survey (D-survey) is a collection of the form $\{Q_n\}_{n\in\mathcal{N}}$, with $Q_n \subseteq \mathcal{Q}$, describing the set of questions that are endorsed by individual n.[6] We study whether the survey data can be rationalized.

D-survey rationalizability. We say that $\{Q_n\}_{n\in\mathcal{N}}$ is rationalizable whenever there exist $[\{U_n, \tau_n\}_{n\in\mathcal{N}}, \{\mu_q\}_{q\in\mathcal{Q}}]$, with $U_n : \mathbb{R} \to \mathbb{R}$ being strictly quasi-concave and $\tau_n, \mu_q \in \mathbb{R}$, such that for every $n \in \mathcal{N}$ and $q \in \mathcal{Q}$, it is $q \in Q_n$ if and only if $U_n(\mu_q) \geq \tau_n$.

In this notion of rationalization, the real line describes all possible views on a topic. Each individual has a strictly quasi-concave utility function, U_n, and a threshold parameter, τ_n. Then, each question occupies a specific position in the real line, μ_q, and is endorsed by individual n whenever the utility of the view represented by this question is above her threshold, $U_n(\mu_q) \geq \tau_n$. Notice that strict quasi-concavity guarantees that each individual has an ideal point, that we denote by μ_n, and that the upper contour set determined by τ_n corresponds to an interval around μ_n.

3.1. **Expanding the information: overlapping intervals.**

We start by arguing that we can expand the informational content of a D-survey. We do so by including the responses of fictitious individuals that, under rationalizability, are implied by

[5]Note that one cannot use the standard results of completion of binary relations. We need to show that our ternary relation can be completed by using techniques from the c1p literature.

[6]We are assuming that all individuals respond to all questions. However, the same analysis goes through when there is missing data, perhaps due to attrition or to loss of data or by design, and the information of each individual corresponds to a subset of questions.

the responses in the D-survey. The trick exploits the basic properties of overlapping intervals of the real line.

Suppose that we observe two response sets Q_{n_1} and Q_{n_2} that are overlapping, i.e., $Q_{n_1} \cap Q_{n_2}$, $Q_{n_1} \setminus Q_{n_2}$ and $Q_{n_2} \setminus Q_{n_1}$ are non-empty. For rationalizability to hold, we know that Q_{n_1} corresponds to the questions in the upper-contour set of U_{n_1} determined by τ_{n_1}, which is an interval of the real line. The same logic applies to Q_{n_2} and, since the response sets of individuals n_1 and n_2 overlap, it follows that their underlying intervals must also overlap. Now, since the intersection and differences of any two overlapping intervals are also intervals, we can understand them as the upper contour sets of some fictitious individuals.[7] Hence, we can expand the informational content in the D-survey by including responses $Q_{n_1} \cup Q_{n_2}$, $Q_{n_1} \setminus Q_{n_2}$ and $Q_{n_2} \setminus Q_{n_1}$. This method can be used recursively to end up with what we call the expanded D-survey, denoted by $\{\bar{Q}_m\}_{m=1}^{M}$, which is the unique collection of responses satisfying: (i) it contains $\{Q_n\}_{n\in\mathcal{N}}$, (ii) it is closed under union and differences of overlapping sets, and (iii) it is the minimal collection satisfying (i) and (ii).

Example 1. Six individuals are inquired about five questions. As it will be discussed in the proof of Theorem 1, for the purposes of rationalizability it suffices to entertain response sets that are all different, contain at least two endorsed questions, and differ from $\mathcal{Q}$. Suppose that we observe the following responses: $Q_1 = \{1,3\}$, $Q_2 = \{2,4,5\}$, $Q_3 = \{3,5\}$, $Q_4 = \{1,2,3,5\}$, $Q_5 = \{2,4\}$ and $Q_6 = \{1,3,5\}$. Individual 1 overlaps with individual 3 but no new information can be obtained from combining their responses because $Q_1 \cup Q_3 = Q_5$ and the differences of their responses are singletons. Individual 2 overlaps with individual 3 and this pair provides a novel response set through the union of individual responses. No other novel pattern can be obtained through union or difference of overlapping sets and hence $M = 7$ with $\bar{Q}_m = Q_m$, $m < 7$, and $\bar{Q}_7 = Q_2 \cup Q_3 = \{2,3,4,5\}$. □

3.2. **Exploiting the information: intermediate questions.** We are ready to exploit the information contained in the expanded D-survey. The key insight of our analysis is that all relevant information is ordinal and it can be represented by a linear order over Q describing the relative position of questions in the real line. We obtain

[7]Notice that this statement does not necessarily apply if the responses are not overlapping: if two intervals are disjoint, the union may fail to be an interval; if one interval contains the other, the difference may fail to be an interval. Also, notice that the intersection of overlapping intervals must also be an interval, but our revelation techniques do not require the addition of this information.

this information by combining the responses of overlapping individuals and using the fact that a linear order can be constructed from its *intermediateness* relation.

Consider two individuals $\{m_a, m_b\}$ whose response sets overlap. Then, the interval nature of their response sets reveals that for any three questions $\{q_a, q_I, q_b\}$ such that $q_a \in \bar{Q}_{m_a} \setminus \bar{Q}_{m_b}$, $q_I \in \bar{Q}_{m_a} \cap \bar{Q}_{m_b}$ and $q_b \in \bar{Q}_{m_b} \setminus \bar{Q}_{m_a}$, q_I must be the one that occupies an intermediate position in the real line. When this pattern is observed, we say that q_I has been revealed intermediate in the triplet $\{q_a, q_I, q_b\}$ and simply write $q_I \in I(\{q_a, q_I, q_b\})$.

Example 1 (continued). As an illustration of how intermediate revelations are produced, notice that the response set $\bar{Q}_1 = \{1, 3\}$ overlaps with the response set $\bar{Q}_3 = \{3, 5\}$, and hence it must be that question 3 is intermediate in the triplet $\{1, 3, 5\}$. The consideration of all pairs of overlapping individuals brings the following revealed intermediate relations: $3 \in I(\{1, 2, 3\})$, $2 \in I(\{1, 2, 4\})$, $5 \in I(\{1, 2, 5\})$, $3 \in I(\{1, 3, 4\})$, $3 \in I(\{1, 3, 5\})$, $5 \in I(\{1, 4, 5\})$, $2 \in I(\{2, 3, 4\})$, $5 \in I(\{2, 3, 5\})$, $2 \in I(\{2, 4, 5\})$, $5 \in I(\{3, 4, 5\})\}$. $\square$

3.3. **Characterization.** Rationalizability merely requires the revealed intermediate relation to satisfy the following acyclicity type of condition.

Strong Axiom of Revealed Intermediateness (SARI). If $q_t \in I(\{q_{t-1}, q_t, q_{t+1}\})$ for every $t = 1, \dots T$, then $q_0, q_T \notin I(\{q_0, q_{T-1}, q_T\})$.

Theorem 1. *$\{Q_n\}_{n \in \mathcal{N}}$ is rationalizable if and only if $\{\bar{Q}_m\}_{m=1}^{M}$ satisfies SARI.*

Proof of Theorem 1: We start by noticing that we can assume without loss of generality that $q \neq q'$ implies $N_q \neq N_{q'}$, where $N_q \subseteq \mathcal{N}$ denotes the set of individuals that endorse question $q \in \mathcal{Q}$. From the *if* point of view, this assumption is innocuous because the same location can be assigned to all questions such that $N_q = N_{q'}$. From the *only if* point of view, this assumption is innocuous because the revelations produced by two questions satisfying $N_q = N_{q'}$ are fully symmetric, and the acyclicity described by SARI is not affected by the elimination of one of these questions.

Also, notice that we can assume without loss of generality that all response sets in $\{\bar{Q}_m\}_{m=1}^{M}$ are different, i.e., $m \neq m'$ implies $\bar{Q}_m \neq \bar{Q}_{m'}$. From the *if* point of view, this assumption is innocuous because all individuals such that $\bar{Q}_m = \bar{Q}_{m'}$ can be represented by the same utility and threshold. From the *only if* point of view, this assumption is innocuous because two individuals such that $\bar{Q}_m = \bar{Q}_{m'}$ produce the same expansions and revelations.

Finally, we can also assume without loss of generality that all response sets in $\{\bar{Q}_m\}_{m=1}^M$ contain at least two questions, but are not equal to $\mathcal{Q}$. From the *if* point of view, this assumption is innocuous because: (i) any individual that endorses all questions can be added by selecting any U_m and setting a sufficiently low τ_m, (ii) any individual that endorses a unique question q can be added by selecting any U_m and τ_m such that $\mu_m = \mu_q$ and $U_m(\mu_q) = \tau_m$, and (iii) any individual that endorses no question can be added by selecting any U_m and setting a sufficiently high τ_m. From the *only if* point of view, this assumption is innocuous because any such response set does not produce any revelation, either directly or via expansion.

We now proceed to prove the result by means of a series of claims.

Claim 1. $\{Q_n\}_{n\in\mathcal{N}}$ is rationalizable if and only if $\{Q_n\}_{n\in\mathcal{N}}$ satisfies the consecutive ones property (or c1p), i.e., there exists a linear order $\succ$ over $\mathcal{Q}$ such that, for every individual $n \in \mathcal{N}$, Q_n is formed by consecutive questions in $\succ$.

Proof of Claim 1: First, suppose that the D-survey is rationalizable. Then, it must be so for a set of parameters $[\{U_n, \tau_n\}_{n\in\mathcal{N}}, \{\mu_q\}_{q\in\mathcal{Q}}]$. For every pair of questions $q, q' \in \mathcal{Q}$, since they are not endorsed by the same subset of individuals, it must be $\mu_q \neq \mu_{q'}$. Hence, we can define a linear order $\succ$ by setting $q \succ q' \Leftrightarrow \mu_q > \mu_{q'}$. Now, for every $n \in \mathcal{N}$, rationalizability guarantees that $Q_n = \{q \in \mathcal{Q} : U_n(\mu_q) \geq \tau_n\}$ and given that U_n is strictly quasi-concave, the questions in Q_n must be consecutive in $\succ$.

Suppose now that the survey is c1p. Then, it must be so for some linear order $\succ$. Given the finiteness of $\mathcal{Q}$, we know that this linear order is representable and, hence, there exists a collection of real values $\{\mu_q\}_{q\in\mathcal{Q}}$ such that for every pair of questions $q, q' \in \mathcal{Q}$, it is $\mu_q > \mu_{q'} \Leftrightarrow q \succ q'$. For every $n \in \mathcal{N}$, define $\mu_n = \frac{\min_{q\in Q_n} \mu_q + \max_{q\in Q_n} \mu_q}{2}$ and set $U_n(\mu_n) = 1$. Since $|Q_n| \geq 2$, it must be $\min_{q\in Q_n} \mu_q < \mu_n < \max_{q\in Q_n} \mu_q$. We can set $U_n(\min_{q\in Q_n} \mu_q) = U_n(\max_{q\in Q_n} \mu_q) = 0$ and complete U_n to be the piecewise linear function fully determined by these three points.

Finally, define $\tau_n = 0$. The function U_n is strictly quasiconcave. Since Q_n is formed by consecutive questions in $\succ$, any question q' in Q_n, and only these, must satisfy $\max_{q\in Q_n} \mu_q \geq \mu_{q'} \geq \min_{q\in Q_n} \mu_q$, or equivalently $U_n(\mu_{q'}) \geq \tau_n$. Rationalization has been proved. $\square$

Claim 2. $\{Q_n\}_{n\in\mathcal{N}}$ satisfies c1p if and only if $\{\bar{Q}_m\}_{m=1}^M$ satisfies c1p.

Proof of Claim 2: First suppose that $\{Q_n\}_{n\in\mathcal{N}}$ satisfies c1p. Then, there exists a linear order $\succ$ such that, for every $n \in \mathcal{N}$, Q_n is formed by a set of consecutive questions in $\succ$. Any union and difference of two overlapping sets of consecutive questions in $\succ$ is also a consecutive set in $\succ$. Since every response set in $\{\bar{Q}_m\}_{m=1}^M$ can be obtained by a finite recursion of this principle, the linear order $\succ$ also guarantees that the expanded D-survey satisfies c1p.

Suppose now that $\{\bar{Q}_m\}_{m=1}^M$ satisfies c1p. Since the D-survey is contained in the expanded D-survey, any linear order for which the expanded D-survey satisfies c1p also guarantees that the D-survey does so. This concludes the proof of the claim. $\square$

Claim 3. $\{\bar{Q}_m\}_{m=1}^M$ satisfies c1p if and only if $\{\bar{Q}_m\}_{m=1}^M$ satisfies SARI.

Proof of Claim 3: Suppose first that $\{\bar{Q}_m\}_{m=1}^M$ satisfies c1p. Then, there exists a linear order $\succ$ such that, for every $1 \leq m \leq M$, $\bar{Q}_m$ is formed by a set of consecutive questions in $\succ$. Consider a sequence of questions $(q_0, q_1, \ldots, q_T)$ such that $q_t \in I(\{q_{t-1}, q_t, q_{t+1}\})$ for every $t = 1, \ldots, T$. Since $q_1 \in I(\{q_0, q_1, q_2\})$, there must exist a pair of individuals endorsing question q_1, with one of them endorsing q_0 but not q_2 and the other endorsing q_2 but not q_0. Given that the response sets of these two individuals are consecutive in $\succ$, it must be either $q_0 \succ q_1 \succ q_2$ or $q_2 \succ q_1 \succ q_0$. Similarly, since $q_2 \in I(\{q_1, q_2, q_3\})$, it must be either $q_1 \succ q_2 \succ q_3$ or $q_3 \succ q_2 \succ q_1$. Moreover, these two facts can be combined to learn that the quadruple must be ordered as either $q_0 \succ q_1 \succ q_2 \succ q_3$ or $q_3 \succ q_2 \succ q_1 \succ q_0$. The iterative application of this principle shows that it is either $q_0 \succ \cdots \succ q_T$ or $q_T \succ \cdots \succ q_0$. In any case, since all response sets in $\{\bar{Q}_m\}_{m=1}^M$ are formed by consecutive questions in $\succ$, $\{q_0, q_T\} \subseteq \bar{Q}_m$ implies $q_{T-1} \in \bar{Q}_m$ and hence neither q_0 nor q_T can be revealed intermediate in the triplet $\{q_0, q_{T-1}, q_T\}$, as desired.

Suppose now that $\{\bar{Q}_m\}_{m=1}^M$ satisfies SARI. We proceed by contradiction and assume that $\{\bar{Q}_m\}_{m=1}^M$ fails to satisfy c1p. Then, its incidence matrix must fail the matrix version of c1p, and it must contain one sub-matrix adopting one out of the five forbidden configurations found by Tucker (1972; Theorem 9).[8] We analyze each one separately

[8]By incidence matrix we refer to the matrix assigning a value of one to the cell defined by row m and column q whenever $q \in Q_m$, and a zero otherwise. The definition of c1p for matrices is given in Section 2.

using our set notation. In each case, the subsets of individuals and questions involved in the sub-matrix are denoted by $\mathcal{M}^*$ and $\mathcal{Q}^*$, with cardinalities M^* and Q^*.[9]

Case I. $M^* = Q^* \geq 3$ where, for every $m < M^*$, $\bar{Q}_m \cap \mathcal{Q}^* = \{m, m+1\}$ and $\bar{Q}_{M^*} \cap \mathcal{Q}^* = \{1, M^*\}$. When this is the case, it is immediate to see that for every $m < M^* - 1$, the pair of overlapping individuals $\{m, m+1\}$ reveals that $m + 1 \in I(\{m, m+1, m+2\})$. Moreover, the pair of overlapping individuals $\{1, M^*\}$ reveals that $1 \in I(\{1, 2, M^*\})$. The conjunction of all these revelations is a direct violation of SARI, and a contradiction has been reached.

Case II. $M^* = Q^* \geq 4$ where, for every $m < M^* - 1$, $\bar{Q}_m \cap \mathcal{Q}^* = \{m, m+1\}$, $\bar{Q}_{M^*-1} \cap \mathcal{Q}^* = \{1, 2, \ldots, M^* - 2, M^*\}$ and $\bar{Q}_{M^*} \cap \mathcal{Q}^* = \{2, 3, \ldots, M^*\}$. When this is the case, sets $\bar{Q}_1$ and $\bar{Q}_2$ overlap, and hence, there must exist an individual with response set equal to $\bar{Q}_1 \cup \bar{Q}_2$. When constrained to $\mathcal{Q}^*$, this response set is equal to $\{1, 2, 3\}$. We can use the same logic over the latter individual and $\bar{Q}_3$, and continue doing so recursively. Ultimately, we can find an individual i such that $\bar{Q}_i \cap \mathcal{Q}^* = \{1, 2, \ldots, M^* - 2\}$. Now, notice that $\bar{Q}_i$ and $\bar{Q}_{M^*}$ are overlapping sets and hence, there must exist an individual, that we denote by m_{M^*-1}, such that $\bar{Q}_{m_{M^*-1}} = \bar{Q}_{M^*} \setminus \bar{Q}_i$ and hence, it must be $\bar{Q}_{m_{M^*-1}} \cap \mathcal{Q}^* = \{2, 3, \ldots, M^*\} \setminus \{1, 2, \ldots, M^* - 2\} = \{M^* - 1, M^*\}$. Using the same recursive logic than above, we can consider all individuals from 2 to $M^* - 2$ and identify an individual j such that $\bar{Q}_j \cap \mathcal{Q}^* = \{2, \ldots, M^* - 1\}$. Since this response set overlaps with $\bar{Q}_{M^*-1}$, the difference of $\bar{Q}_{M^*-1}$ and $\bar{Q}_j$ must also correspond to an individual, that we denote by m_{M^*}. It is $\bar{Q}_{m_{M^*}} \cap \mathcal{Q}^* = \{1, M^*\}$. It suffices to notice that the pattern of responses of individuals $\{1, 2, \ldots, M^* - 2, m_{M^*-1}, m_{M^*}\}$ over Q^* has exactly the structure described in Case I, reaching a contradiction.

Case III. $M^* = Q^* - 1 \geq 3$ where, for every $m < M^*$, $\bar{Q}_m \cap \mathcal{Q}^* = \{m, m+1\}$, $\bar{Q}_{M^*} \cap \mathcal{Q}^* = \{2, 3, \ldots, M^* - 1, M^* + 1\}$. Using the same logic as in Case II, we can recursively consider the first $M^* - 2$ individuals and find a response set such that, when constrained to $\mathcal{Q}^*$, is equal to $\{1, 2, \ldots, M^* - 1\}$. Individuals $M^* - 1$ and M^* are overlapping, and hence we can consider their union to obtain a response set that, when constrained to $\mathcal{Q}^*$, is equal to $\{2, 3, \ldots, M^* + 1\}$. These two constructed response sets are overlapping and hence, there must be an individual, that we denote by m_{M^*}, such that $\bar{Q}_{m_{M^*}} \cap \mathcal{Q}^* = \{2, 3, \ldots, M^* + 1\} \setminus \{1, 2, \ldots, M^* - 1\} = \{M^*, M^* + 1\}$. In a similar fashion, we can consider the union of the response sets of individuals

[9]Since in the arguments that follow the order and labeling of such individuals and questions are irrelevant, we adopt the most convenient notation.

M^* and 1 on one hand, and the union of the response sets of individuals from 2 up to M^*-1 on the other. The difference between these two overlapping response sets must correspond to an individual, which we denote by m_{M^*+1}, with a response set such that $\bar{Q}_{m_{M^*+1}} \cap \mathcal{Q}^* = \{1, M^*+1\}$. Then, the response sets of individuals $\{1, 2, \ldots, M^*-1, m_{M^*}, m_{M^*+1}\}$ over $\mathcal{Q}^*$ have again the structure of Case I, reaching a contradiction.

Case IV. $M^* = 4$ and $Q^* = 6$, with $\bar{Q}_1 \cap \mathcal{Q}^* = \{1,2\}$, $\bar{Q}_2 \cap \mathcal{Q}^* = \{3,4\}$, $\bar{Q}_3 \cap \mathcal{Q}^* = \{5,6\}$ and $\bar{Q}_4 \cap \mathcal{Q}^* = \{2,4,6\}$. Individual 4 overlaps with all the other three individuals, and we can then consider individuals m_j, $j \in \{1,2,3\}$ such that $\bar{Q}_{m_j} = \bar{Q}_4 \setminus \bar{Q}_j$. The response sets of individuals m_1, m_2 and m_3 over the questions $2, 4$ and 6 again have the cyclical structure of Case I, reaching a contradiction.

Case V. $M^* = 4$ and $Q^* = 5$, with $\bar{Q}_1 \cap \mathcal{Q}^* = \{1,2\}$, $\bar{Q}_2 \cap \mathcal{Q}^* = \{1,2,3,4\}$, $\bar{Q}_3 \cap \mathcal{Q}^* = \{3,4\}$ and $\bar{Q}_4 \cap \mathcal{Q}^* = \{1,4,5\}$. Individual 4 overlaps with the other three, and hence we can identify individuals m_j, $j \in \{1,2,3\}$ such that $\bar{Q}_{m_1} = \bar{Q}_4 \setminus \bar{Q}_1$, $\bar{Q}_{m_2} = \bar{Q}_2 \setminus \bar{Q}_4$ and $\bar{Q}_{m_3} = \bar{Q}_4 \setminus \bar{Q}_3$. The response sets of individuals $1, m_2, 3, m_1$ and m_3 over $\mathcal{Q}^*$ have again the structure of Case I, which is a contradiction and concludes the proof of the claim. $\square$

Hence, $\{Q_n\}_{n\in\mathcal{N}}$ is rationalizable if and only if $\{Q_n\}_{n\in\mathcal{N}}$ satisfies c1p, if and only if $\{\bar{Q}_m\}_{m=1}^M$ satisfies c1p, if and only if $\{\bar{Q}_m\}_{m=1}^M$ satisfies SARI, and the result has been proved. $\blacksquare$

The strategy of the proof is as follows. First, we argue that the model is ordinal and show that rationalizability is equivalent to the construction of an order of questions $\succ$ such that the endorsements of each individual form an interval, i.e., we need to show that the survey responses satisfy c1p. Second, we use the fact that the union and differences of overlapping intervals of the real line are also intervals and show that the original survey satisfies c1p if and only if the expanded survey does so. Finally, we show that the expanded survey satisfies c1p if and only if it satisfies SARI. We do so using a classical result by Tucker (1972) in which all possible violation patterns of c1p are obtained. When working with the expanded survey, we show that each of these patterns can be mapped into a cyclical violation of SARI. Importantly, the use of expanded surveys is key here; the acyclicity of revelations in the original survey is not sufficient for rationalizability. We illustrate this by means of the following simple example.

Example 2. Three individuals are inquired about four questions, producing response sets $Q_1 = \{1,2\}$, $Q_2 = \{1,3\}$ and $Q_3 = \{1,4\}$. Every pair of individuals is overlapping, and the revealed relation is $1 \in I(\{1,2,3\}), 1 \in I(\{1,2,4\}), 1 \in I(\{1,3,4\})$. SARI is vacuously satisfied. However, the expanded D-survey fails SARI because it contains all possible subsets of two or three questions. Hence, no rationalization is possible. □

Theorem 1 shows that the model is ordinal. In the proof, we set the parameters as follows. We argue that we can locate the questions using any real values $\{\mu_q\}_{q\in\mathcal{Q}}$ that represent $\succ$. Then, for every individual n, $U_n(\mu_q)$ is set to be the negative absolute value distance between the middle point of the interval of the locations of the endorsed questions and that of question q. The threshold τ_n is defined by the utility of the first and last questions in the interval of questions endorsed by n.

4. Rationalizability of attitudinal surveys: extensions

In this section, we show how to exploit the structure of Theorem 1 when dealing with two related, informationally richer, problems. The first one covers the case in which individuals are allowed to use more than two labels in declaring their endorsement. The second analyzes the case of dichotomous surveys in which responses may be probabilistic.

4.1. Polytomous surveys.

Given question $q \in \mathcal{Q}$, individual $n \in \mathcal{N}$ is now allowed to express strength of endorsement by using a label from the collection $\mathcal{L} = \{0, 1, \ldots, L\}$, where $L > 0$. We will write $\mathcal{L}^* = \mathcal{L}\backslash\{0\}$. Higher labels in the collection express stronger endorsement.[10] A polytomous survey or, simply, a P-survey is a map $P : \mathcal{N} \times \mathcal{Q} \to \mathcal{L}$, where $P(n,q) = l$ refers to the case where individual n assigns label l to question q. We consider the following notion of rationalizability, which reduces to the notion used for D-surveys when $L = 1$.

P-survey rationalizability. We say that P is rationalizable whenever there exist $[\{U_n, \{\tau_n^l\}_{l\in\mathcal{L}^*}\}_{n\in\mathcal{N}}, \{\mu_q\}_{q\in\mathcal{Q}}]$, with $U_n : \mathbb{R} \to \mathbb{R}$ being strictly quasi-concave, $\tau_n^1 < \cdots < \tau_n^L$, and $\mu_q \in \mathbb{R}$ such that, for every $n \in \mathcal{N}$, $q \in \mathcal{Q}$ and $l \in \mathcal{L}^*$, $P(n,q) = L$ if $U_n(\mu_q) \geq \tau_n^L$, $P(n,q) = l$, $0 < l < L$, if $\tau_n^{l+1} > U_n(\mu_q) \geq \tau_n^l$ and $P(n,q) = 0$ otherwise.

In words, the individual now uses different thresholds to determine the degree of endorsement. These thresholds naturally impose stronger requirements on stronger expressions of endorsement.

[10] For instance, extremely disagree, disagree, neutral, agree, extremely agree.

We now show that the rationalizability of a P-survey can be obtained building upon Theorem 1, by studying the rationalizability of an associated D-survey. Formally, given the response map P, define the D-survey $\{Q^P_{(n,l)}\}_{(n,l)\in\mathcal{N}\times\mathcal{L}^*}$ where $Q^P_{(n,l)} = \{q \in \mathcal{Q} : P(n,q) \geq l\}$. That is, the vector of responses of individual $n \in \mathcal{N}$ is used to construct $L-1$ response sets. Each of these response sets contains cumulative information, e.g., response set $Q^P_{(n,l)}$ contains all the questions endorsed by individual n with intensity l or above. Note that, formally, $\{Q^P_{(n,l)}\}_{(n,l)\in\mathcal{N}\times\mathcal{L}^*}$ is a D-survey where every (n,l) can be understood as an individual. Now, it is immediate that $\{Q^P_{(n,l)}\}_{(n,l)\in\mathcal{N}\times\mathcal{L}^*}$ provides the same information as P does. We need to define the corresponding expanded D-survey, that follows the same logic than above, and that we denote by $\{\bar{Q}^P_{(m)}\}_{m=1}^M$. Theorem 2 shows that the acyclicity of the intermediateness revelation of $\{\bar{Q}^P_{(m)}\}_{m=1}^M$ is equivalent to the rationalizability of P.

Theorem 2. *P is rationalizable if and only if $\{\bar{Q}^P_{(m)}\}_{m=1}^M$ satisfies SARI.*

Proof of Theorem 2: We start by assuming that no questions are symmetric and not two individuals assign the same labels to all questions. Analogous arguments to those in Theorem 1 shows that this is without loss of generality.

First suppose that P is rationalizable. Then, it must be so for some collection of parameters $[\{U_n, \{\tau^l_n\}_{l\in\mathcal{L}^*}\}_{n\in\mathcal{N}}, \{\mu_q\}_{q\in\mathcal{Q}}]$. Define $q \succ q'$ if and only if $\mu_q > \mu_{q'}$. It is immediate that any response set $Q^P_{(n,l)}$ must be formed by consecutive questions in $\succ$. Hence, the D-survey $\{Q^P_{(n,l)}\}_{(n,l)\in\mathcal{N}\times\mathcal{L}^*}$ satisfies c1p and using the only if parts of Claims 2 and 3 in the proof of Theorem 1, $\{\bar{Q}^P_{(m)}\}_{m=1}^M$ must satisfy SARI.

Suppose now that $\{\bar{Q}^P_{(m)}\}_{m=1}^M$ satisfies SARI. Using the if part of Claim 3 in Theorem 1, $\{\bar{Q}^P_{(m)}\}_{m=1}^M$ must satisfy c1p. Let $\succ$ be a linear order that guarantees so and consider any set of values $\{\mu_q\}_{q\in\mathcal{Q}}$ such that $\mu_q > \mu_{q'} \Leftrightarrow q \succ q'$. Given individual $n \in \mathcal{N}$, let $l_n > 0$ be the highest label used by this individual. If this label is used to describe endorsement for at least two questions, define $\mu_n = \frac{\min_{q\in Q_{(n,l_n)}} \mu_q + \max_{q\in Q_{(n,l_n)}} \mu_q}{2}$ and set $U_n(\mu_n) = l_n$ and $U_n(\min_{q\in Q_{(n,l_n)}} \mu_q) = U_n(\max_{q\in Q_{(n,l_n)}} \mu_q) = l_n - 1$. All values in the interval $[\min_{q\in Q_{(n,l_n)}}, \max_{q\in Q_{(n,l_n)}}]$ are set by considering the piecewise linear function determined by these three values. If the label was used for only one question, set μ_n equal to the μ-value of that question, and $U_n(\mu_n) = l_n - 1$. After this, consider recursively the rest of the labels, starting with $l_n - 1$ down to label 1. For any label l, determine the questions $\min_{q\in Q_{(n,l)}} \mu_q$ and $\max_{q\in Q_{(n,l)}} \mu_q$. If any of them is different to those previously considered (i.e., whenever these questions have label l exactly), extend the piecewise linear function U_n by forcing it to adopt value $l-1$ in the location of these new questions. After label 1, extend the function beyond its extremes with any strictly decreasing linear function. Set thresholds $\tau_n^l = l - 1$. The function U_n is strictly quasi-concave, and since $Q_{(n,l)}$ is formed by a set of consecutive questions in $\succ$, it is immediate that questions in $Q_{(n,l)}$, and only these questions, have utility above threshold τ_n^l. Rationalization then follows. ■

4.2. **Stochastic responses.** We now study the case where survey responses may be stochastic. This can be understood in terms of the variation of responses of the individuals of a given subgroup, like those given by different ages or genders, or questionnaires asking for the expression of probabilistic beliefs, as it is often the case in experiments, or the variation of responses of an individual across repetitions of a questionnaire. Stochastic models of responses are instrumental in estimation exercises and in accounting for errors and other behavioral considerations such as fatigue or inattention.

Formally, a dichotomous survey with stochastic responses, or S-survey, is a map $S : \mathcal{N} \times \mathcal{Q} \to [0,1]$ where $S(n,q)$ describes the probability with which individual $n \in \mathcal{N}$ endorses question $q \in \mathcal{Q}$. Consider the following definition of rationalizability.

S-survey rationalizability. We say that S is rationalizable whenever there exist $[\{U_n, F_n\}_{n\in\mathcal{N}}, \{\mu_q\}_{q\in\mathcal{Q}}]$, with $U_n : \mathbb{R} \to \mathbb{R}$ being strictly quasi-concave, $F_n : \mathbb{R} \to [0,1]$ being a cumulative distribution function (CDF) governing the realization of thresholds τ_n, and $\mu_q \in \mathbb{R}$ such that, for every $n \in \mathcal{N}$ and $q \in \mathcal{Q}$, $S(n,q) = F_n(U_n(\mu_q))$.

Notice that any question is endorsed whenever the utility of the view represented by this question is above a threshold but now, these thresholds are realized according to a distribution F_n. Hence, endorsement happens probabilistically, and the probability of individual n endorsing question q corresponds to the mass of thresholds below $U_n(\mu_q)$, i.e., $F_n(U_n(\mu_q))$. This is a convenient account of stochastic responses, that sits well with the models studied in this paper.

We now argue that we can rely again on the techniques developed for the analysis of D-surveys. For every $n \in \mathcal{N}$, consider the $|\mathcal{Q}|$ response sets defined by $Q_n^q = \{q' \in \mathcal{Q} : S(n,q') \geq S(n,q)\}$ and the corresponding D-survey $\{Q^S_{(n,q)}\}_{(n,q)\in\mathcal{N}\times\mathcal{Q}}$. That is, the information of individual $n \in \mathcal{N}$ appears $|\mathcal{Q}|$ times, describing response sets that contain the collection of questions chosen with at least as much probability as a given question. Denote the corresponding expanded D-survey by $\{\bar{Q}^S_{(m)}\}_{m=1}^M$. Theorem 3 shows that the rationalizability of S follows from the acyclicity of this expanded survey.

Theorem 3. *S is rationalizable if and only if $\{\bar{Q}^S_{(m)}\}_{m=1}^M$ satisfies SARI.*

Proof of Theorem 3: Since the proof is very similar to that of Theorem 2, we only mention the most significant details. Notice that we can construct a hypothetical set of $Q \times N$ labels, and define the P-survey with the response map P having the property that, for every $n, n' \in \mathcal{N}$ and $q, q' \in \mathcal{Q}$, $P(n,q) \geq P(n',q')$ if and only if $S(n,q) \geq S(n',q')$. This allows us to construct the values μ_q and the functions U_n exactly as in Theorem 2. For every τ_n such that there exists $q \in \mathcal{Q}$ with $U_n(\mu_q) = \tau_n$, set $F_n(\tau_n) = S(n,q)$. Now, for every individual $n \in \mathcal{N}$, a larger threshold corresponds to a question with a larger utility and given the construction and the assumption that $P(n,q) \geq P(n,q')$ if and only if $S(n,q) \geq S(n,q')$, a larger utility corresponds to a larger probability of endorsement. We can then extend these values to a CDF F_n over the reals, and rationalization follows. ■

5. Identification

In this section, we discuss the identification of the models studied. As discussed above, these models are essentially ordinal, so we start asking ourselves under what conditions the linear order $\succ$ on the set of questions is fully identified. Later, we analyze what we can learn about the ideal point of any individual. Finally, to allow for cardinal identification results, a parametric version of the model is introduced and analyzed.

To present results across different types of surveys, we make the following assumptions. We assume that for every individual $n \in \mathcal{N}$: (i) $\emptyset \neq Q_n \neq \mathcal{Q}$, (ii) $S(n,q) = S(n,q') \Rightarrow P(n,q) = P(n,q') \Rightarrow [q \in Q_n$ if and only $q' \in Q_n]$. That is, (i) implies that we eliminate from the analysis the trivial individuals that endorse everything or nothing, and (ii) connects the response behavior in S-surveys with that of P-surveys and of D-surveys.

We impose the following richness condition. To present it, consider for concreteness the dichotomous case. In essence, for a triplet of questions being relatively close to each other, one can expect that there are two individuals revealing its intermediateness information. However, when questions are further apart, this assumption may be too strong to impose, and accordingly we postulate a milder, indirect, version. Formally, for every triplet of distinct questions T, there exists a sequence of questions $(q_1, q_2, \ldots, q_K)$ such that: (i) T is contained in the sequence, and (ii) for every $k \in \{1, 2, \ldots, K-2\}$, there exists two individuals $n_{k1}, n_{k2} \in \mathcal{N}$ such that $Q_{n_{k1}} \cap \{q_k, q_{k+1}, q_{k+2}\} = \{q_k, q_{k+1}\}$ and $Q_{n_{k2}} \cap \{q_k, q_{k+1}, q_{k+2}\} = \{q_{k+1}, q_{k+2}\}$.

5.1. **Identification of questions.** We start by showing that $\succ$ is fully identified. The reason for this is that, under the richness assumption, the map I can be shown to be complete. As a result, it identifies the question with an intermediate μ-value for every triplet, and the linear order is fully identified (up to inversion).[11]

Proposition 1. *If a survey is rationalizable, it is for a unique linear order $\succ$ (up to inversion). This linear order is fully revealed by the intermediateness relation I.*

Proof of Proposition 1: Let the survey be rationalizable. Hence, the intermediateness relation I of the expanded survey satisfies the acyclicity type of condition postulated by SARI. We now show that I is also complete, in the sense that it reveals at least one intermediate question in every triplet. Consider any triplet of distinct questions T. By our richness condition, there exists a sequence of questions $(q_1, q_2, \ldots, q_K)$ such that: (i) T is contained in the sequence, and (ii) for every $k \in \{1, 2, \ldots, K-2\}$, there exist $n_{k1}, n_{k2} \in \mathcal{N}$ such that $Q_{n_{k1}} \cap \{q_k, q_{k+1}, q_{k+2}\} = \{q_k, q_{k+1}\}$ and $Q_{n_{k2}} \cap \{q_k, q_{k+1}, q_{k+2}\} = \{q_{k+1}, q_{k+2}\}$. That is, two real individuals have revealed that $q_{k+1} \in I(\{q_k, q_{k+1}, q_{k+2}\})$. We now claim that for every triplet of different questions in this sequence satisfying $k' < k'' < k'''$, it must be $q_{k''} \in I(\{q_{k'}, q_{k''}, q_{k'''}\})$.[12] We prove this claim recursively, over the difference $k''' - k'$. The result is trivially true for $k''' - k' = 2$ because in this case, the triplet must be formed by three questions that appear consecutively in the sequence. Suppose now that the result is true up to some value $k''' - k' = t \geq 2$ and consider a triplet of questions such that $k''' - k' = t + 1$. Let q_* be any question in the sequence occupying a position between $q_{k'}$ and $q_{k'''}$ and such that $q_* \neq q_{k''}$.[13] Assume that $q_{k''}$ comes before q_* in the sequence (the other case is dual and thus omitted). Given the recursive step, we know that $q_{k''} \in I(\{q_{k'}, q_{k''}, q_*\})$ and $q_* \in I(\{q_{k''}, q_*, q_{k'''}\})$. Therefore, there exist four individuals $m_1, m_2, m_3, m_4 \in \{1, \ldots, M\}$ such that $\bar{Q}_{m_1} \cap \{q_{k'}, q_{k''}, q_*\} = \{q_{k'}, q_{k''}\}$, $\bar{Q}_{m_2} \cap \{q_{k'}, q_{k''}, q_*\} = \{q_{k''}, q_*\}$,

[11] Since this result holds for all types of surveys, we do not specify the type in the formulation of the result. Also, we simplify the exposition of the proof by showing the key steps using only the D-survey notation.

[12] Notice that these will be revelations that can be the result of real or hypothetical individuals in the expanded survey.

[13] One such question must exist because $t + 1 \geq 3$.

$\bar{Q}_{m_3} \cap \{q_{k''}, q_*, q_{k'''}\} = \{q_{k''}, q_*\}$, $\bar{Q}_{m_4} \cap \{q_{k''}, q_*, q_{k'''}\} = \{q_*, q_{k'''}\}$. Given rationalizability and the fact that $q_* \in I(\{q_{k'}, q_*, q_{k'''}\})$, it must obviously be the case that $q_{k'''} \notin \bar{Q}_{m_1}$. If $q_{k'''} \in \bar{Q}_{m_2}$, individuals m_1 and m_2 reveal that $q_{k''} \in I(\{q_{k'}, q_{k''}, q_{k'''}\})$, as desired. Otherwise, individuals m_2 and m_4 overlap, and since the expanded survey is closed by unions, there exists an individual m_5 such that $\bar{Q}_{m_5} = \bar{Q}_{m_2} \cup \bar{Q}_{m_4}$. It is $\bar{Q}_{m_5} \cap \{q_{k'}, q_{k''}, q_{k'''}\} = \{q_{k''}, q_{k'''}\}$. This individual overlaps with m_1 and reveals that $q_{k''} \in I(\{q_{k'}, q_{k''}, q_{k'''}\})$, as desired. Since the triplet is contained in the sequence and the argument applies to any triplet, this shows that I is complete. As a result, there must be a unique order (up to inversion), captured by this intermediateness relation.[14] Moreover, given the revelation analysis, this order must correspond to the one induced by $\{\mu_q\}$ (or its inverse), concluding the proof. ■

Example 1 (continued). There are no cycles in the intermediateness relation I. For instance, notice that $3 \in I(\{1,2,3\})$ and $2 \in I(\{2,3,4\})$ require that neither 4 nor 1 can be revealed intermediate in $\{1,2,4\}$, as it is the case. Hence, the expanded D-survey satisfies SARI and the D-survey is rationalizable. Moreover, I is complete and hence there is a unique (up to inversion) linear order of the questions compatible with the intermediate relation, $1 \succ 3 \succ 5 \succ 2 \succ 4$.

Notice that no pair of real individuals reveals information on the triplet $\{q_1, q_2, q_3\}$. The intermediateness information for this triplet can be obtained throughout the sequence (q_1, q_3, q_5, q_2). Two real individuals (1 and 3) reveal that q_3 is the intermediate question in the triplet $\{q_1, q_3, q_5\}$. Two real individuals (2 and 3) reveal that q_5 is the intermediate question in the triplet $\{q_3, q_5, q_2\}$. The information of these individuals can be linked to produce the hypothetical individual 7 (union of 2 and 3), allowing to conclude that $q_3 \in I(\{q_1, q_2, q_3\})$. □

5.2. **Identification of individuals.** We now discuss what can be potentially learned about a specific individual $n \in \mathcal{N}$, concentrating our efforts in the ordinal identification of her ideal point. We show that S-surveys are more informative in this respect than P-surveys, which in turn are more informative than D-surveys. For any individual $n \in \mathcal{N}$, denote by $\underline{q}_n^\alpha$ and $\overline{q}_n^\alpha$, with $\alpha \in \{\text{D,P,S}\}$, the minimum and maximum questions (according to μ_q or $\succ$) among: (i) the endorsed ones when $\alpha = \text{D}$, (ii) those with a higher label than any other when $\alpha = \text{P}$, and (iii) those with higher probability than

[14] It can also be constructed simply Take any three questions and determine the intermediate one, locating the other two as desired. Proceed including questions one by one, in the unique position allowed by I.

the rest when $\alpha = \text{S}$. Then, denote by $\alpha(n)$ the open interval of $\succ$ defined by the question immediately below $\underline{q}_n^\alpha$ and the one immediately above $\overline{q}_n^\alpha$ in $\succ$.[15] Importantly, in S-surveys whenever the CDF over the threshold is atomless, these two questions must either coincide or be consecutive, implying that the identification is almost complete.[16]

Proposition 2. *The ideal view of individual $n \in \mathcal{N}$ belongs to $S(n) \subseteq P(n) \subseteq D(n)$, and no further identification is possible.*

Proof of Proposition 2: Let the survey be rationalizable and $\succ$ be the unique linear order guaranteeing so. Consider $n \in \mathcal{N}$ and an α-survey. First, we claim that the ideal view must belong to the interval $\alpha(n)$. Suppose by way of contradiction that this is not the case. We analyze the case in which the ideal view occupies a position equal or above the question immediately above $\overline{q}_n^\alpha$, and omit the other, dual case. In this case, notice that the question $\overline{q}_n^\alpha$ achieves a higher endorsement than the question above, contradicting the quasi-concavity of U_n. Second, we claim that $S(n) \subseteq P(n) \subseteq D(n)$. To see this, consider first $q \in S(n)$. By construction, q achieves the highest endorsement probability in $\mathcal{Q}$ and, given our basic assumptions, there exists at least one question q' with strictly less endorsement probability. Question q must belong to the equivalence class of questions with the highest label in the P-survey, and this level must be different than zero. It then follows that $S(n) \subseteq P(n)$. Consider now a question $q \in P(n)$. By construction, q achieves the highest label in $\mathcal{Q}$ and, given our basic assumptions, this label is strictly above zero. Question q must belong to the response set in the D-survey and $P(n) \subseteq D(n)$. Third, we claim that no further identification is possible. We analyze the case of D-surveys, with the other two being analogous. Consider the set Q_n. Modify the selection of the ideal point in the proof of Theorem 1 to occupy any position in $D(n)$. If this position is between $\overline{q}_n^\alpha$ and $\underline{q}_n^\alpha$, set the utility function to be piecewise linear, with value 1 for the ideal view and 0 for these two questions. Otherwise, it must be either above $\overline{q}_n^\alpha$ or below $\underline{q}_n^\alpha$. Consider the first case (the second being omitted as it is analogous). Then, set the utility function to be piecewise linear, with value 1 for the ideal view, 0 for $\underline{q}_n^\alpha$, and -1 for the question immediately above $\underline{q}_n^\alpha$

[15]If one of these two questions does not exist, the position of the ideal view can be seen as unbounded in that specific direction.

[16]This follows immediately from the fact that, given strict quasi-concavity and the atomless of the CDF, only the closest question at each side of the ideal point may come with the largest endorsement probability (up to copies of these questions).

(if no question exists there, any strictly decreasing function works). This alternative representation also provides rationalizability and completes the proof. ■

5.3. **Exponential responses.** In the models discussed so far, the identification of questions and ideal points is ordinal. We now introduce a stochastic parametric model that provides cardinal identification of the location of questions and ideal points, $\{\mu_q\}_{q\in\mathcal{Q}}$ and $\{\mu_n\}_{n\in\mathcal{N}}$. The model postulates that the individual endorsement probability is of the exponential form on the distance between the locations of the question and the ideal point of the individual, $S(n,q)=e^{-\frac{|\mu_n-\mu_q|}{\sigma_n}}$.[17] Intuitively, the probability of endorsing a question is 1 when its location coincides with that of the ideal point, decays with the distance between them, and is zero in the limit. It is modulated by a single parameter $\sigma_n>0$ with a natural interpretation. Smaller values of σ_n accelerate the decay, representing individuals with a less tendency to endorse views away from their ideal point. Notice that the exponential endorsement probability could be alternatively presented via our stochastic framework by way of a strictly quasi-concave utility and a random threshold.[18]

We start by normalizing the location of any two questions, that we denote by 0_q and 1_q, to $\mu_{0_q}=0$ and $\mu_{1_q}=1$, respectively.[19] We further assume that there exist at least two individuals that place the largest probability of endorsement in different questions and that there are at least four questions.

Proposition 3. *In an S-survey with exponential responses, $\{\mu_q\}_{q\in\mathcal{Q}\setminus\{0_q,1_q\}}$ and $\{\mu_n\}_{n\in\mathcal{N}}$ are cardinally identified.*

Proof of Proposition 3: Consider an S-survey with exponential responses. By Proposition 1, we know that $\succ$ is fully identified, and we have set it to be $1_q\succ 0_q$. Given the atomless nature of the exponential model, for every $n\in\mathcal{N}$, the interval $S(n)$ contains either one or two questions. As in the proof of Proposition 2, denote these two questions by $\overline{q}_n^S$ and $\underline{q}_n^S$, and it must be either $\overline{q}_n^S\succ\underline{q}_n^S$ or $\overline{q}_n^S=\underline{q}_n^S\equiv q_n^S$. We now construct a map $h_n:\mathcal{Q}\to\{0,1\}$ for each individual such that $h_n(q)=1$ if and only if $\mu_q\geq\mu_n$ and $h_n(q)=0$ otherwise. To do so, we consider two cases.

[17] The literature in psychology offers various probabilistic models of responses based on the notion of endorsement-by-proximity. See, e.g., Davison (1977). The present model is a convenient version of Andrich (1988) and Hoijtink (1990).

[18] It suffices to set $U_n(x)=e^{-\frac{|\mu_n-x|}{\sigma_n}}$ and F_n the CDF of the uniform distribution in $[0,1]$.

[19] Recall that the inverse of a linear order is equivalent to the linear order for our purposes, so the order of these two alternatives is without loss of generality.

If $|S(n)| = 2$, the exponential model of responses guarantees that μ_n must occupy a position strictly between these two questions, and we can assign $h_n(q) = 1$ to $\overline{q}_n^S$ and to any other question above this one, according to $\succ$, and zero otherwise. If $|S(n)| = 1$, the map can be defined by $h_n(q) = 1$ whenever $q \succ q_n^S$ and $h_n(q) = 0$ whenever $q_n^S \succ q$. However, the relative location of q_n^S is yet undetermined. To do so, we can simply consider another individual n' for which $h_{n'}(q_n^S)$ is known, and two questions q', q'' for which $h_n(q'), h_n(q''), h_{n'}(q'), h_{n'}(q'')$ are all known. It is immediate to see that the exponential model requires that

$$\frac{(-1)^{h_n(q_n^S)} \log S(n, q_n^S) - (-1)^{h_n(q'')} \log S(n, q'')}{(-1)^{h_n(q')} \log S(n, q') - (-1)^{h_n(q'')} \log S(n, q'')} = \frac{\mu_{q_n^S} - \mu_{q''}}{\mu_{q'} - \mu_{q''}} =$$

$$= \frac{(-1)^{h_{n'}(q_n^S)} \log S(n', q_n^S) - (-1)^{h_{n'}(q'')} \log S(n', q'')}{(-1)^{h_{n'}(q')} \log S(n', q') - (-1)^{h_{n'}(q'')} \log S(n', q'')}.$$

Hence, $h_n(q_n^S)$ can be determined.[20] Given the complete map h_n, notice now that for every q other than 0_q and 1_q it is

$$\frac{(-1)^{h_n(q)} \log S(n, q) - (-1)^{h_n(0_q)} \log S(n, 0_q)}{(-1)^{h_n(1_q)} \log S(n, 1_q) - (-1)^{h_n(0_q)} \log S(n, 0_q)} = \mu_q,$$

$$\frac{-(-1)^{h_n(0_q)} \log S(n, 0_q)}{(-1)^{h_n(1_q)} \log S(n, 1_q) - (-1)^{h_n(0_q)} \log S(n, 0_q)} = \mu_n.$$

All locations are fully determined, concluding the proof. ■

6. Rationalizability of aptitude surveys

We now briefly study the rationalizability of surveys measuring aptitudes, rather than attitudes. These are surveys with the aim of studying absolute, cumulative, aptitudes, as it is often the case in the study of ability (say, the mathematical skills of pupils), health status (say, the extent of mobility for the aged), social functionality (say, the social skills of psychiatric patients), or the extent of a practice (say, the level of

[20]Notice that in the extreme case in which $S(n, q_n^S) = 1$, it must be $\mu_n = \mu_{q_n^S}$ and $h_n(q_n^S)$ can be freely assigned.

religiosity of citizens). Here we focus on the dichotomous case, which is usually known as Guttman-scale.[21] For obvious reasons, we refer to it as Guttman-rationalizability.

Guttman-rationalizability. We say that $\{Q_n\}_{n\in\mathcal{N}}$ is Guttman-rationalizable whenever there exist $[\{\mu_n\}_{n\in\mathcal{N}}, \{\mu_q\}_{q\in\mathcal{Q}}]$ with $\mu_n, \mu_q \in \mathbb{R}$, such that for every $n \in \mathcal{N}$ and $q \in \mathcal{Q}$, it is $q \in Q_n$ if and only if $\mu_n \geq \mu_q$.

In a Guttman-rationalizability, a question (or problem, or task) q is passed whenever its location falls below the location of the individual, $\mu_q \leq \mu_n$. Notice that Guttman-rationalizability is a particular case of D-survey rationalizability since we could alternatively rationalize responses by constructing a strictly quasi-concave utility function U_n and a threshold τ_n such that the individual endorses every question located below the ideal point, and only these.[22] Hence, SARI is a necessary condition for Guttman-rationalizability, but not sufficient. We now provide a necessary and sufficient condition. Importantly, the cumulative nature of Guttman-rationalizability facilitates the survey revelation exercise. It allows us: (i) to analyze the responses by way of a binary relation approach, instead of the ternary approach we adopted above, and (ii) to use only the information directly contained in the survey, without requiring the expansion of the survey data.

Consider a pair of questions $q_1, q_2 \in \mathcal{Q}$ and an individual $n \in \mathcal{N}$ such that $Q_n \cap \{q_1, q_2\} = q_2$. It is then evident that the aptitude of individual n separates the complexity of both questions, and the complexity of question q_1 must be above that of question q_2. We then say that q_1 has been revealed to be more complex than question q_2, and simply write $q_1 \succ_{mc} q_2$. A standard acyclicity condition on the binary relation $\succ_{mc}$ is enough for Guttman-rationalizability.

Strong Axiom of Revealed Complexity (SARC). If $q_t \succ_{mc} q_{t-1}$ for every $t = 1, \ldots, T$, then $q_0 \not\succ_{mc} q_T$.

Theorem 4. *$\{Q_n\}_{n\in\mathcal{N}}$ is Guttman-rationalizable if and only if $\{Q_n\}_{n\in\mathcal{N}}$ satisfies SARC.*

[21]The extensions to non-dichotomous and probabilistic responses follow analogously to the cases studied above.

[22]For instance, consider any utility function such that $\lim_{x\to-\infty} U_n(x) = 1$, increasing up to $U_n(\mu_n) > 1$, and decreasing afterwards with $U_n(\mu_n + \epsilon) < 1$ for any sufficiently small $\epsilon > 0$ such that there is no question in the interval $(\mu_n, \mu_n + \epsilon)$, and $\tau_n = 1$. This configuration gives positive answers until the ideal point of the individual, and negative ones afterwards.

Proof of Theorem 4: Suppose first that $\{Q_n\}_{n\in\mathcal{N}}$ is Guttman-rationalizable. If $q_t \succ_{mc} q_{t-1}$ for every $t = 1, \ldots, T$, we know that there exist individuals n_t, $t = 1, \ldots, T$ such that $\mu_{q_T} > \mu_{n_T} \geq \mu_{q_{T-1}} > \mu_{n_{T-1}} \geq \mu_{q_{T-2}} > \cdots \geq \mu_{q_1} > \mu_{n_1} \geq \mu_{q_0}$ and hence, $\mu_{q_T} > \mu_{q_0}$. Then, it is evident that $q_0 \not\succ_{mc} q_T$.

Suppose now that $\{Q_n\}_{n\in\mathcal{N}}$ satisfies SARC. Given the finiteness of $\mathcal{Q}$, we can find values $\{\mu_q\}_{q\in\mathcal{Q}}$ such that $q \succ_{mc} q'$ implies $\mu_q > \mu_{q'}$. Now, if an individual responds negatively to all questions set $\mu_n < \min_{q\in\mathcal{Q}} \mu_q$. Otherwise, set $\mu_n = \max_{q\in Q_n} \mu_q$. We claim that this constitutes a valid Guttman-rationalization. Suppose, by way of contradiction, that this is not the case. Notice that the construction guarantees that every question such that $\mu_q > \mu_n$ is negatively responded. The contradiction must involve an individual $n \in \mathcal{N}$ and a question $q \in \mathcal{Q}$ such that $\mu_n \geq \mu_q$ but $q \notin Q_n$. However, our construction guarantees that n responds positively to at least one question, and the definition of μ_n guarantees that there is a question q' with $\mu_{q'} > \mu_q$ and $q' \in Q_n$. However, this would make n to reveal $q \succ_{mc} q'$, which leads to $\mu_q > \mu_{q'}$, a contradiction. This concludes the proof. ■

7. Discussion

This paper represents a first attempt in the study of the rationality foundations of survey responses, setting the basis for the theoretical treatment of other important issues.[23] For example, our treatment uses the classical unidimensional approach, in which questions and ideal points can be located in a common scale. One may naturally wonder about the multidimensional case. Whenever each question belongs to one known dimension, our results follow immediately. This may be the case, for instance, in surveys on political attitudes and national identity. In this case, it is sufficient to analyze the survey data separately for each dimension, using exactly the techniques we have developed in this paper. In general, the treatment of cases where questions belong to more than one dimension may be challenging, requiring techniques other than the ones we have developed in this paper. We leave this for further research, but note here that if one allows for any number of dimensions, the survey data can always be rationalizable. The intuition would be as follows. Locate every question in a different dimension, and set the ideal point of an individual as a convex combination of

[23]Another direction of interest involves the combination of surveys with more standard economic data to enhance the information revelation process, in line with the recent suggestions of Caplin (2021) and Almås, Attanasio, and Jervis (2023).

the questions she endorses. The Euclidean distance of any endorsed question is always smaller than that of any other question.

References

[1] Acemoglu, D., Makhdoumi, A., Malekian, A., and Ozdaglar, A. (2022). Learning from reviews: The selection effect and the speed of learning. *Econometrica*, 90(6), 2857-2899.

[2] Almås, I., Attanasio, O., and Jervis, P. (2023). Economics and Measurement: New measures to model decision making (No. w30839). National Bureau of Economic Research.

[3] Alós-Ferrer, C. and Buckenmaier, J. (2019). Strongly sincere best responses under approval voting and arbitrary preferences. *Games and Economic Behavior*, 117, 388-401.

[4] Andrich, D. (1988). The application of an unfolding model of the PIRT type to the measurement of attitude. *Applied Psychological Measurement*, 12(1), 33-51.

[5] Barberá, P. (2014). Birds of the same feather tweet together: Bayesian ideal point estimation using Twitter data. *Political Analysis*, 23(1), 76-91.

[6] Benjamin, D. J., Guzman, J. D., Fleurbaey, M., Heffetz, O., and Kimball, M. S. (2021). What do happiness data mean? Theory and survey evidence (No. w28438). National Bureau of Economic Research.

[7] Bertrand, M. and Mullainathan, S. (2001). Do people mean what they say? Implications for subjective survey data. *American Economic Review*, 91(2), 67-72.

[8] Caplin, A. (2021). Economic Data Engineering (No. w29378). National Bureau of Economic Research.

[9] Coombs, C. H. (1964). *A theory of data*. Wiley.

[10] Falk, A., Neuber, T., and Strack, P. (2021). Limited self-knowledge and survey response behavior. Mimeo.

[11] Fishburn, P. C. (1971). Betweenness, orders and interval graphs. *Journal of Pure and Applied Algebra*, 1(2), 159-178.

[12] Guttman, L. (1944). A basis for scaling qualitative data. *American Sociological Review*, 9(2), 139-150.

[13] Greene, W. H. and Hensher, D. A. (2010). *Modeling ordered choices: A primer*. Cambridge University Press.

[14] Hoijtink, H. (1990). A latent trait model for dichotomous choice data. *Psychometrika*, 55, 641-656.

[15] Huntington, E. V. and Kline, J. R. (1917). Sets of independent postulates for betweenness. *Transactions of the American Mathematical Society*, 18(3), 301-325.

[16] Hubert, L. (1974). Problems of seriation using a subject by item response matrix. *Psychological Bulletin*, 81(12), 976.

[17] Kendall, D. G. (1969). Some problems and methods in statistical archaeology. *World Archaeology*, 1(1), 68-76.

[18] Laslier, J. F. (2009). The leader rule: A model of strategic approval voting in a large electorate. *Journal of Theoretical Politics*, 21(1), 113-136.

[19] Liiv, I. (2010). Seriation and matrix reordering methods: An historical overview. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 3(2), 70-91.

[20] Manzini, M., Mariotti, M, and Ülkü, L. (2022). A model of approval. Mimeo.

[21] Núñez, M. and Xefteris, D. (2017). Implementation via approval mechanisms. *Journal of Economic Theory*, 170, 169-181.

[22] Poole, K. T. and Rosenthal, H. (2007). *Congress: A political-economic history of roll call voting*. Oxford University Press.

[23] Rasch, G. (1961). On general laws and the meaning of measurement. In *Psychology, Proceedings of the Fourth Berkley Symposium on Mathematical Statistics and Probability*, University of California Press, 5, 321-333.

[24] Stantcheva, S. (2022). How to run surveys: A guide to creating your own identifying variation and revealing the invisible. *Annual Review of Economics*, 15.

[25] Tay, L. and Ng, V. (2018). Ideal point modeling of non-cognitive constructs: Review and recommendations for research. *Frontiers in Psychology*, 2423.

[26] Thurstone, L. L. (1928). Attitudes can be measured. *American Journal of Sociology*, 33(4), 529-554.

[27] Tucker, A. (1972). A structure theorem for the consecutive 1's property. *Journal of Combinatorial Theory*, Series B, 12(2), 153-162.

[28] Van der Linden, W. J. (2018). *Handbook of item response theory: Three volume set*. CRC Press.