FREEDOM COUNTS: CROSS-COUNTRY EMPIRICAL EVIDENCE*

João V. Ferreira
Nobuyuki Hanaki
Fabrice Le Lec
Erik Schokkaert
Benoît Tarroux†

November 7, 2023

Abstract

This paper investigates how people evaluate different sets of opportunities in terms of *welfare* and *freedom of choice*. To do this, we run a new survey-based study with 4,902 participants across 10 different countries, in which subjects face a series of theoretically-relevant binary comparisons of *opportunity sets*. Our analysis proceeds in two stages. We first use a naive Bayesian method to classify subjects according to the theoretical rules they implicitly employ to compare sets in terms of freedom and welfare. Then, we investigate whether subjects value freedom of choice even if more freedom does not lead to the choice of a better alternative (*intrinsic value of freedom of choice*). Our main result is that an overwhelming majority of subjects reveal attaching intrinsic value to freedom. We also find that a large majority of subjects use size-based rules to rank sets in terms of freedom, while there is considerable heterogeneity in the theoretical rules they employ to rank sets in terms of welfare. These results are strikingly robust across countries. All this suggests that it is important to offer choice to individuals in the design of organizations and public policies, even if this does not substantially change their choice behavior.

Keywords: Freedom of choice; Welfare; Intrinsic value; Opportunity set; Cross-cultural survey.

JEL Codes: C83, D63, I31.

---

*This paper benefits from the funding of the French National Agency for Research research project (The Value of Choice, Grant No. ANR-16-CE41-0002-01), the Japan Society for the Promotion of Science (JSPS), and the Joint Usage/Research Center at ISER, Osaka University.

†*Corresponding author*. University Lumière Lyon 2 and GATE, France. E-mail: benoit.tarroux@univ-lyon2.fr; Postal address: 93, Chemin des Mouilles, 69130 Ecully, France; Phone: +670009022.

I. INTRODUCTION

How to rank *opportunity sets* — the set of options available to individuals — is an important normative and empirical question, with implications for organizations, markets, and public policy. In economics, it is standard practice to rank opportunity sets based on the "best" alternatives available on those sets, where "best" is defined on the basis of individual preferences. According to this approach (called *indirect utility*, or IU for short), freedom of choice (hereafter FoC) has only *instrumental value*: additional alternatives in an opportunity set are valuable only if they lead to the choice of a better alternative in the individual's preference ranking. Although this approach is standard and parsimonious, two important empirical questions remain unanswered.

First, do people only attach instrumental value to freedom, or do they also attach *intrinsic value* to it? The literature on FoC — as pioneered by Amartya Sen (1985, 1988) and developed by many others — postulated that additional alternatives can be valuable even if they do *not* lead to the choice of a better alternative.[1] Why would individuals value freedom beyond its (purely) instrumental benefits? Possible underlying reasons are that FoC, independently of the alternatives that are chosen, is an element of a person's well-being (e.g. Sen 1988), is important to develop mental faculties such as judgment and self-control (e.g. Mill 1859, p. 117), or allows people to lead autonomous and therefore meaningful lives (e.g. Nozick 1974, pp. 48-51; Arrow 1995). However, to the best of our knowledge, there is little if any empirical evidence on whether people attach intrinsic value to freedom.

A second related and important question, for which there is also very limited evidence, is to know how people compare different opportunity sets in terms of freedom and welfare.[2] Which theoretical ranking rules (like IU) do people (implicitly) employ

[1] See Sugden (1998), Barbera et al. (2004), Baujard (2007), Gravel (2008b), Dowding and van Hees (2009), and Foster (2011) for reviews. The literature on the *capability approach* also highlights both the instrumental and intrinsic values of freedom (see Basu and López-Calva 2011 and Robeyns 2017 for a review).

[2] "Welfare" is here understood as an all-things-considered evaluation of sets (see Hausman 2012). In other words, everything that contributes to the value of an opportunity set is taken into account for welfare.

when ranking opportunity sets? This is again an empirical question. It is distinct from the normative question — which has received much attention in the economic literature (see e.g. Barbera et al. 2004 and Foster 2011 for reviews) — of how we *should* rank opportunity sets. However, following the tradition of *positive welfare economics*, we argue that it can inform the latter question (see e.g. Gaertner and Schokkaert 2012 and Ambuehl and Bernheim 2021). For example, understanding how people rank opportunity sets can bring valuable insights into the political feasibility and democratic desirability of different theoretical ranking rules.

In this paper, we use a novel survey-based research design to i) understand whether people attach intrinsic value to FoC, and ii) investigate which theoretical rules people implicitly employ when ranking opportunity sets in terms of FoC and overall welfare. We do this for a total of 4902 participants across 10 different countries.[3] Participants face 15 comparisons of opportunity sets in a health-related context. For each *set comparison*, we ask them to evaluate which set provides more FoC (our Freedom question), what is the "best" alternative present in the sets (our IU question), and which set is best overall (our Welfare question). To investigate whether people attach intrinsic value to FoC, we look at set comparisons where subjects respond that there is a "conflict" between FoC and the "best" alternative (i.e., when a set A has more FoC and the best alternative is in a set B or in both sets).[4] Then, if a participant states that set A is best overall, we say that this participant *reveals* attaching intrinsic value to freedom. To investigate which rules people use, we draw on the theoretical literature and identify a reasonably large number of plausible theoretical rules that can be used to rank sets in terms of FoC and welfare. Each of these rules implies a distinctive *theoretical response pattern* for the 15 set comparisons. We then compare this theoretical response pattern to the *actual response pattern* of participants using a Bayes classification procedure, in order to assign participants to

[3]Our list of countries includes Brazil, China, Colombia, France, Japan, Netherlands, Portugal, Turkey, the United Kingdom, and the US.

[4]These are the only relevant set comparisons for this research question since, by construction, a participant cannot reveal attaching intrinsic value to freedom when set A has more FoC and it has a better alternative than all available in set B.

the theoretical rule that best matches their responses in the Freedom question and the (potentially different) rule that best matches their responses in the Welfare question.

We find that an overwhelming majority of subjects ($> 80\%$) *reveal* that they attach intrinsic value to freedom at least once. In addition, participants attach intrinsic value to freedom in 48% of the relevant cases. Regarding our second question, we find, first, that the overwhelming majority of subjects respond *as if* they rely on cardinality-based rules (rules that "count" the number of alternatives) to rank sets in terms of FoC. Second, when ranking sets in terms of overall welfare, we find that quality-based rules (such as IU or other rules that take into consideration the "quality" of alternatives) dominate, but there is considerable heterogeneity in the rules that subjects implicitly employ. In terms of country comparisons, we find that our results are strikingly similar across countries with very different social, cultural, and institutional backgrounds. This suggests that the intrinsic value of freedom and people's rankings of opportunity sets are robust cross-cultural attitudes. Finally, we show that the fit of our classification results is overall good, that we have not omitted empirically important rules, and that our main results for the classification and intrinsic value of freedom are not driven by random answers, mistakes or intuitive responses, salience effects, preferences for flexibility, or other potential confounding factors.

Our analysis is important from both an academic and a policy perspective. From an academic point of view, our analysis contributes to several important literatures in economics. Our paper is related to the literature on positive welfare economics (e.g. Yaari and Bar-Hillel 1984; Gaertner and Schokkaert 2012; Ambuehl and Bernheim 2021). Following the seminal paper by Yaari and Bar-Hillel (1984), there is a large literature on positive welfare economics (or "empirical social choice") using survey-based, lab, and online experiments to test if fundamental principles of social choice and welfare economics, like the Pareto principle, are supported by the general public. Opinions of lay people are seen as relevant empirical facts to determine the plausibility, democratic desirability, and political feasibility of theoretical postulates. We contribute to this literature in several ways. First, we provide the first (to our knowledge) empirical evidence on the theoretical rules that people implicitly employ when ranking opportunity sets and we are the first to study if people attach intrinsic value to freedom. This brings essential insights about the theoretical assumptions of the FoC literature cited above, and suggests that the intrinsic value of freedom should be taken into consideration in the growing literature studying the trade-off between preferences and freedom (e.g. Arad and Rubinstein 2018; Ackfeld and Ockenfels 2021; Ambuehl et al. 2021; Dreyer and Mahler 2022; Alsan et al. 2023). Second, our results for 10 countries offer crucial insights into the evaluation of opportunity sets for citizens from a large variety of social, cultural, and institutional backgrounds. In particular, our results uncover attitudes concerning freedom that seem to be robust across countries. Finally, from a methodological point of view, our analysis of the intrinsic value of freedom provides a clear empirical test that can be used in other studies.

Our paper also contributes to the growing empirical literature on choice over menus. Following seminal theoretical contributions to the study of preferences over menus (e.g. Kreps 1979; Gul and Pesendorfer 2001), several lab experimental studies have sought to elicit these preferences (e.g. Toussaert 2018; Le Lec and Tarroux 2020; Arlegi et al. 2022). For example, Toussaert (2018) uses choices between menus to elicit the cost of self-control in the presence of temptation (as theorized by Gul and Pesendorfer 2001), and finds that more than a quarter of subjects in the lab prefer to remove tempting options from their opportunity sets. In a different perspective, Le Lec and Tarroux's (2020) lab experiment provides evidence that subjects value less a menu than its preferred option, a phenomenon that can be rationalized by subjects' fear of making mistakes in the presence of a larger menu. Yet, these approaches do not directly test concerns for freedom: the fact that subjects choose menus themselves implies that, overall, their freedom is not affected. Other lab experiments have shown that subjects are willing to pay for the intrinsic value of choosing for themselves when interacting with others can reduce their decision rights (e.g. Bartling et al. 2014; Ferreira et al. 2020). However, like the previous experiments, these studies do not allow to measure in detail concerns for freedom of choice. As a consequence, we contribute to this growing literature by measuring directly whether freedom is an important concern, as well as the extent to which it influences preference over menus. Overall, our results tend to show that the intrinsic value of freedom of choice substantially influences preferences over menus and that IU will disregard relevant welfare-enhancing features of opportunity sets (see Benjamin et al. 2014 for related evidence).

From a policy perspective, the attitudes of the population on social issues can be an important input into the process of democratic public decision-making. If (as we find) a large share of the population attaches intrinsic value to freedom, this implies that people's preferences over menus can complement information about preferences over alternatives as inputs to public policy. If policymakers wish to respect people's preferences over menus, our results can also inform them on how to compare different opportunity sets. For instance, our results strongly suggest that for a substantial part of the population, it is important to enlarge small opportunity sets even if this does not lead to a change in behavior. This is an important finding since many policy decisions are between providing a single option or a small number of options, as is the case in many countries for the choice of schools, hospitals, or pension schemes in companies.[5]

The remainder of the paper is organized as follows. In the next section, we formally define what we mean by the intrinsic value of freedom of choice and we present the theoretical ranking rules that we test in our data. In Section III, we present the research design of the study. Section IV is devoted to the empirical analysis and results, and Section V concludes.

[5]Our study is set in the hospital choice context and brings direct evidence to this setting. Our findings may be influenced to some extent by this specific context. However, we believe that the questions in our survey are sufficiently abstract to downplay the possible importance of this specific context. In addition, the application to the health domain is interesting in its own right, as it counts for a large percentage of GDP across OECD countries and in some countries (like France, Portugal, and the UK in our sample) introducing more personal choice in the health care sector is an important ongoing policy debate (see e.g. UK Government 2023).

II. CONCEPTUAL BACKGROUND

To illustrate the problem, consider the following simple example:

- **Opportunity set A:** (apple, orange)
- **Opportunity set B:** (apple)

where "apple" and "orange" are possible alternatives for consumption and opportunity sets A and B are menus from which those fruits can be chosen. In this section, we formally define what it means for an individual to reveal attaching intrinsic value to FoC in such situations, and we summarize the main rules that have been proposed in the theoretical literature to rank different opportunity sets in terms of freedom and welfare.

Before proceeding, we introduce a bit of notation. Let X be a finite set of alternatives, denoted below by x, y, and z, and let A and B denote non-empty subsets of X (i.e., opportunity sets). Let $\succsim^F$ be a transitive quasi-ordering over opportunity sets such that $A \succsim^F B$ means that "A provides at least as much freedom of choice as B", with $\sim^F$ and $\succ^F$ being the symmetric and asymmetric components of $\succsim^F$. We denote by $\succsim^W$ a transitive quasi-ordering over opportunity sets such that $A \succsim^W B$ means that "A provides at least as much welfare as B", with $\sim^W$ and $\succ^W$ being the symmetric and asymmetric components of $\succsim^W$. Our interpretation of this welfare relation is an *all-things-considered* comparison of sets. Let $\succsim_i^W$ and $\succsim_i^F$ correspond to $\succsim^W$ and $\succsim^F$ as judged by individual i. Finally, let R_i denote individual i's transitive and complete preference ordering over feasible alternatives (with P_i and I_i its asymmetric and symmetric components) and $\max(A)$ denote individual i's set of preferred element(s) of A such that $\max(A) = \{x : xR_iy \; \forall y \in A\}$. With a slight abuse of notation, we write $\max(A) R_i \max(B)$ to denote that individual i weakly prefers the best element(s) in A to those in B, with $\max(A) P_i \max(B)$ for strict preference and $\max(A) I_i \max(B)$ for indifference.

II.A The intrinsic value of freedom of choice

How can an individual *reveal* that he/she attaches intrinsic value to freedom of choice (hereafter also IvFoC) through their choices? Formally:

Definition 1. ***(IvFoC).*** *For two sets A and B such that $A \succ_i^F B$ and $\max(B) R_i \max(A)$, individual i reveals attaching IvFoC if $A \succ_i^W B$.*

In other words, we say that i reveals attaching IvFoC whenever faced with sets A and B such that i judges that (i) A provides more FoC than B and (ii) the best alternative in B is at least as good as the best alternative in A, individual i considers that A provides more overall welfare than B. For example, take the opportunity sets $A = \{x, y, z\}$ and $B = \{x, y\}$. Assume that i considers that set A provides more FoC than set B. If i deems that, all things considered, set A provides more welfare than set B even though the best alternative in both sets is x (i.e., $\max(B) I_i \max(A)$), then i responds *as if* he/she attaches intrinsic value to having more FoC. Now consider $A = \{x, y, z\}$ and $B = \{w\}$. A stronger case of IvFoC is if i considers that A provides more welfare and FoC than set B, even though w is better than x, y, and z (i.e., $\max(B) P_i \max(A)$). Are there individuals who give so much intrinsic value to freedom that they prefer opportunity sets with more freedom but with a "worse" top alternative?

II.B Ranking opportunity sets

The theoretical literature on ranking opportunity sets is vast. Nonetheless, it is possible to organize the existing theoretical ranking rules in different *families*. In this paper, we focus on three main families: the *cardinality family*, focusing on the "size" of the menu (Section II.B.1), the *indirect utility family*, ranking the sets based on the perceived "quality" of their best alternatives (Section II.B.2), and the *potential preferences family*, using multiple "reasonable" preference orderings for the evaluation of sets (Section II.B.3). In Section II.B.4, we discuss some other theoretical rules that we test in the data.

We cover rules that have been proposed to rank sets in terms of FoC, welfare, or both.

Note, however, that all rules can *a priori* be used to rank sets in terms of FoC or welfare. In our empirical analysis, we will test the significance of all the rules we cover in both dimensions. Overall, we test a reasonably large set of plausible ranking rules.[6]

II.B.1 Cardinality family

The first family of rules focuses on the size of the opportunity sets. A seminal ranking rule, the *cardinality rule*, ranks opportunity sets based on the number of alternatives of each set (Pattanaik and Xu 1990). This rule focuses on the "quantity of action" available to a person (Carter 1999) and may be "a natural way of measuring freedom in the absence of information about the agent's preferences" (Foster 2011, p. 20).[7]

The cardinality rule can be weakened to the so-called *weak cardinality rule* (see Puppe 1996 for a related rule). This rule still compares sets based on their size, but states that adding an alternative to a set *will never decrease* (as opposed to *will always increase*) freedom/welfare. As part of this family, we also consider a *diversity rule* that counts the number of "non-similar options" in a set. It says that adding a non-similar option to a set always increases freedom/welfare, while adding an option to a set that is similar to another option available in that set does not increase freedom/welfare.[8]

[6] See Appendix A for a summary of the 23 rules we consider and their formal definitions (all appendixes available online). We show that these rules are exhaustive in our setting: In Appendix D.5, we search for patterns in subjects' responses using a cluster analysis and show that we do not omit important rules in our main analysis. Note that we do not consider *negative freedom* (e.g. Hayek 1960; Van Hees 1998) and other forms of interpersonal freedom (e.g. Sher 2018) that are unimportant (both theoretically and empirically) in our opportunity set context where interpersonal constraints are fixed, but that may affect the evaluation of social states in other settings.

[7] In their seminal paper, Pattanaik and Xu (1990) show how this ranking rule is equivalent to a set of simple axioms. For the sake of concision, we focus on the rules without introducing their equivalent axioms. Doing otherwise would greatly enlarge the conceptual background and divert attention from our main analysis.

[8] There are several competing proposals to measure diversity/similarity (e.g. Pattanaik and Xu 2000; Nehring and Puppe 2002; Bervoets and Gravel 2007). We sidestep this issue in our empirical setting by having options that are clearly similar and options that are clearly dissimilar without reference to preferences.

II.B.2 Indirect utility family

Cardinality-based rules are often criticized for not taking into account the "quality" of the alternatives or the elements present in the sets (e.g. Sen 1990; Sugden 1998). For example, according to cardinality-based rules, a singleton menu A with the option "rotten apple" and a singleton menu B with the option "good apple" should be equally ranked.

An alternative is then to use rules that take the quality of alternatives into consideration. This is particularly relevant if one is ranking opportunity sets in terms of welfare, but quality-based ranking rules have been proposed as freedom rules as well (see e.g. Foster 2011).

Among these, the most prominent is the standard economics approach. Known as the *indirect utility rule*, it ranks sets based on their "best" element, where "best" is defined according to the individual's *actual preference* over alternatives. According to this rule, the menu with the best element for individual i offers greater freedom/welfare.

We also consider quality-based rules that abstract from preferences and define the quality of the elements on a cardinal scale. These rules describe alternatives by their quality in n attributes. The quality of x can then be written as $\sum_j^n w_j x_j$, where x_j is the quality of x in attribute j and w_j is a quality weight that can be varied as a sort of sensitivity analysis. In our empirical setting, alternatives are described by two attributes. Therefore, plausible variations of IU put more or less weight on one of the two attributes. In accordance, we test three *MaxMax rules* that weight the best alternative in one of the two attributes or weight the best alternative giving the same weight to both attributes [i.e., with weights $(w_1, w_2) = (0, 1)$ or $(w_1, w_2) = (1, 0)$ and $(w_1, w_2) = (0.5, 0.5)$ respectively].[9]

We also consider lexicographic rules that combine IU and cardinality, as proposed by Bossert et al. (1994). For example, the *Lex IU-Cardinality rule* first compares sets according to IU, and only in case of indifference based on IU, it then compares sets

[9] We restrict our analysis to this set of weights because there are few ranking differences between these three weights *and* the other potential weights.

according to cardinality. This rule relies on "quality" to make judgments, but when two sets have the same quality level, it uses "quantity" to break the tie. The *Lex Cardinality-IU rule* is the other way around. In our analysis, we attach the *Lex IU-Cardinality rule* to the indirect utility family and the *Lex Cardinality-IU rule* to the cardinality family.

II.B.3 Potential preferences family

The third family we consider appeals to *potential preferences* to take the quality of alternatives into account, where these potential preferences are defined as "the range of preferences that the individual might have had in the relevant circumstances" (Sugden 1998, p. 323). For example, potential preferences can be all preferences that according to the observer can be "reasonably" held by individuals in the context of interest. For example, in most contexts it seems reasonable to prefer to "eat a good apple" than to "eat a good orange", or *vice-versa*; however, it does not seem ever reasonable to prefer to "eat a rotten apple" than to "eat a good apple". In that case, we say that "eat a good apple" and "eat a good orange" are *eligible* options, while "eat a rotten apple" is an *ineligible* option.

A prominent ranking rule in this family is the *range of opportunity rule*, according to which opportunity set A offers more freedom/welfare than B if A caters to more potential preferences than B (Pattanaik and Xu 1998; see also Sugden 1998). More precisely, it says that $A \succsim^F B$ (or $A \succsim^W B$) if and only if the number of *eligible* options in A that are at least as good as all the elements in B according to at least one potential preference is greater than the number of equivalent options in B.

Another ranking rule belonging to the potential preferences family is the *effective freedom rule* put forward by Foster (2011) (see Arrow 1995 and Sen 2002 for related rules). This rule judges one set to have greater or equal freedom/welfare than another set if all potential preferences agree this is so. This ranking rule is "incomplete", as it does not compare sets for which not all potential preferences agree. Finally, we also consider lexicographic rules that combine the range of opportunity and cardinality rules. In our analysis, the *Lex RoO-Cardinality rule* belongs to the potential preferences family and the *Lex Cardinality-RoO rule* belongs to the cardinality family.[10]

II.B.4 Other rules

The three families described in the previous sections contain most of the best-known rules to rank opportunity sets. We now describe some other prominent rules that we test in our data.

First, we look at a *significant choice rule*, according to which opportunity set A offers more freedom/welfare than B if A has more eligible "non-similar" options than B (Pattanaik and Xu 2000; see also Sugden 1998). This rule takes into account both quality (eligibility) and diversity (non-similarity) as defined above. Second, we consider rules that are "opposite" to some of the previous rules. The first of these, the *choice aversion rule*, exhibits a preference for smaller sets when the best alternative is the same. This can be seen as a lexicographic rule that combines IU with "anti-cardinality". The second of these is the *MaxMin rule*, which ranks set A better than B if A has a better "worst" alternative than B [with quality weights $(w_1,w_2) = (0,1)$, $(w_1,w_2) = (1,0)$, and $(w_1,w_2) = (0.5,0.5)$]. The third is a *MaxAverage rule*, which ranks sets according to the average quality of the attributes of all options in those sets [with quality weights $(w_1,w_2) = (0,1)$, $(w_1,w_2) = (1,0)$, and $(w_1,w_2) = (0.5,0.5)$]. The fourth is the *intersection rule* — proposed by Bossert et al. (1994) — that only ranks sets for which IU and cardinality agree: a set offers more freedom/welfare than another set if it offers more freedom/welfare according to both rules. Contrary to the lexicographic rules that try to find a "balance" when the two rules disagree, the intersection rule does not provide a ranking of the sets in those situations (i.e., it is incomplete). Finally, we test a *trivial rule* that states that all opportunity sets offer the same degree of freedom/welfare. This rule is reviewed in Foster (2011), and it is used here mainly as a robustness check.

[10] We do not consider lexicographic rules that combine IU, range of opportunity, and effective freedom, because these are similar quality-based rules.

III. RESEARCH DESIGN

In this section, we present our survey-based research design. Survey-based studies are increasingly popular in economics and, as argued by Stantcheva (2023, pp. 205-6), they allow us to unveil certain attitudes, perceptions, and beliefs that are difficult to reveal through choices. In our study, we present participants with several hypothetical situations with two individuals who face different opportunity sets. We are then interested in how subjects rank the opportunity sets of the two individuals in terms of FoC and welfare. Subjects face 15 *set comparisons* described below.

III.A Context

Participants are presented with a short *vignette* — a brief description of a hypothetical scenario — with a health-related context (see Appendix G for the full instructions). In the vignette, participants are told that two individuals, called Mr. Green and Mr. Yellow, who *are identical in all respects*, have to undergo a surgical procedure that is of minimal risk to their overall health.[11] However, this procedure requires them to spend four days recovering in a hospital and they have to choose a hospital for this surgery and for the recovery time. Importantly, the two individuals have different opportunity sets. Participants are told that their opportunity sets (the hospitals they can choose from) depend on their health insurance plans.[12]

In a typical set comparison, the participants have to compare two sets of hospitals, in which hospitals differ in terms of *staff quality* ("service and assistance quality, nursing quality, friendliness of staff, etc., excluding doctors") and *comfort quality* ("bed quality, food quality, amenities, etc.").[13] Each attribute is rated from 1% for the lowest quality

[11]We called the two individuals Mr. Green and Mr. Yellow to avoid the influence of individual perceptions and local social norms about names.

[12]All countries in our data have some form of private health insurance plan. Therefore, this formulation is a non-artificial reason for opportunity sets to differ across individuals. Still, we do not exclude the possibility that the local organization of hospital care may affect responses.

[13]To avoid staff quality dominating participants' evaluations, we tell them that it excludes doctors and that the hospital choice does not affect Mr. Green or Mr. Yellow's overall health status: "The hospitals available are equivalent in terms of surgery care quality, doctors' skills, etc. Thus, Mr. Green and Mr. Yellow's overall health will not be affected by the choice of hospital for the surgery and the recovering

to 100% for the highest quality (we tell participants that the ratings are from "a trustworthy non-government agency that rates hospitals in their country"). The following set comparison is an example from the survey:

- Mr. Green has the following hospitals available in his insurance plan:
  - Hospital A (staff 80%, comfort 75%)
- Mr. Yellow has the following hospital available in his insurance plan:
  - Hospital A (staff 80%, comfort 75%)
  - Hospital B (staff 71%, comfort 89%)

Before proceeding, it is worth discussing two design choices: (i) why we elicit people's attitudes in a *spectator position*, and (ii) why we describe hospitals with two attributes. First, using a spectator position endows participants with similar information to potential social planners/policymakers.[14] In particular, our participants — like most policymakers — have incomplete information about people's preferences. This is therefore a good position to elicit people's attitudes in the policy-relevant domain. Second, a spectator position allows us to test if people employ prominent rules in the theoretical literature that do not rely on individuals' actual preferences, such as the ranking rules from the potential preferences family. Describing alternatives with two attributes is also important for this aim, since potential preferences rules only differ from IU in set comparisons where there is no dominant alternative, which is made possible by describing alternatives with two attributes. Third, even though we have a spectator position, our design still allows us to test people's attitudes in situations where it is easy to infer Mr. Green and Mr. Yellow's preferences over alternatives. In particular, we use set comparisons where one hospital dominates (or is dominated by) all others in both attributes. As argued below, such situations allow us to exclude some alternative explanations for our

stay."

[14]There is a large and growing literature in economics that uses a spectator position in vignette and experimental studies (e.g. Konow 2009; Almås et al. 2020; Müller and Renes 2021).

results. Finally, describing hospitals with two attributes also allows us to test theoretical ranking rules that take the diversity (or similarity) of alternatives into account.

III.B Main questions

For each set comparison, participants responded to the following three questions (displayed on the same screen without the labels in bold):

- **Q1 (Freedom question):** *Which individual do you think has more freedom of choice?* [Answer options: Mr. Green / Mr. Yellow / The same]
- **Q2 (IU question):** *Which hospital do you think is the best for the treatment and recovery time?* (You can select more than one hospital if you think two or more hospitals are equally best) [Answer options: Hospital A / Hospital B / Hospital C / Hospital D, depending on the hospitals available in the sets]
- **Q3 (Welfare question):** *All things considered, which individual do you think has the best insurance plan?* [Answer options: Mr. Green / Mr. Yellow / Equally good]

Subjects face one of two versions of the survey (between-subject treatments). In version *FreedomIU*, the order of the questions is as presented above, while in version *IUFreedom* the IU question appears first and the freedom question appears second. These orders are kept constant for all set comparisons and the welfare question is always last. The underlying reason for these treatments is two-fold. First, this allows us to test for order effects without increasing participants' cognitive load (as it would be the case, for instance, with a random order of questions for each set comparison). Second, the welfare question is always last as it is framed as an "all things considered" question. While this design choice may make FoC and IU salient in subjects' welfare evaluations, we wanted subjects to consider the welfare question after having compared the two sets in terms of FoC and IU. Doing so allows participants to balance these criteria in their welfare (all-things-considered) evaluations.

III.C The opportunity set comparisons

Participants faced 15 pairwise comparisons of sets presented in a random order. The 15 set comparisons and their rationale are summarized in Table I. Five main reasons underlie the choice of these set comparisons. First, the theoretical response patterns of different ranking rules are distinctive in these set comparisons. As also shown in Table I, this is particularly the case for size-based rules like cardinality and quality-based rules like IU, which we wanted to clearly distinguish. Second, several of these set comparisons are used as "stress tests" of the ranking rules. For example, set comparisons s7 to s9 are increasingly demanding tests of cardinality's implication that adding an alternative always improves FoC/welfare. Third, these set comparisons were chosen such that subjects would likely face set comparisons for which they considered that there is a "conflict" between FoC and the "best" alternative in the sets. As explained above, these are the only situations where a subject can reveal IvFoC. Fourth, having some set comparisons with a clearly dominant or dominated alternative provides us with an in-built test of attention. For example, if subjects state that in s3 the hospital in set B $(60,60)$ is better than the hospital in set A $(80,80)$, then this is a clear mistake. We use this feature of our design to probe our results when we exclude "inattentive" participants. Finally, we focus on small sets with at most three alternatives to limit subjects' cognitive load without restricting our ability to test different theoretical rules, test the intrinsic value of freedom, and bring valuable insights for policy making.

III.D Additional questions

Participants started by stating their gender and age, and at the end of the survey they replied to a short set of questions about their perceived health status, perceived social status, highest level of completed education, occupation, and perceived difficulty of the survey.

	Set comparisons		Underlying reasons for set comparisons	Theoretical response patterns		
	A	*B*		Cardinality	Indirect utility	Range of opportunity
Singleton sets						
s1	$\{(80,80)\}$	$\{(70,90)\}$	Singleton sets without a dominated alternative	∼	A/B/∼	∼
s2	$\{(81,92)\}$	$\{(80,75)\}$	Singleton sets with a dominated alternative	∼	A	A
s3	$\{(80,80)\}$	$\{(60,60)\}$	Singleton sets with a "bad" dominated alternative	∼	A	A
Adding an option						
s4	$\{(80,75)\}$	$\{(80,75);(71,89)\}$	Adding a non-dominated alternative to a singleton	B	B/∼	B
s5	$\{(69,91);(71,89)\}$	$\{(69,91);(71,89);(70,90)\}$	Adding a non-dominated similar alternative	B	B/∼	B
s6	$\{(70,90);(71,89)\}$	$\{(70,90);(71,89);(80,75)\}$	Adding a non-dominated dissimilar alternative	B	B/∼	B
s7	$\{(70,90);(71,89)\}$	$\{(70,90);(71,89);(68,88)\}$	Adding a "slightly" dominated alternative	B	∼	∼
s8	$\{(70,90);(71,89)\}$	$\{(70,90);(71,89);(60,60)\}$	Adding a dominated "bad" alternative	B	∼	∼
s9	$\{(70,90);(71,89)\}$	$\{(70,90);(71,89);(20,30)\}$	Adding a dominated "very bad" alternative	B	∼	∼
s10	$\{(80,75);(95,95)\}$	$\{(80,75);(95,95);(71,89)\}$	Adding an alternative dominated by a "very good" alternative	B	∼	∼
Trade-off between size and quality						
s11	$\{(79,90);(77,91);(80,87)\}$	$\{(81,92)\}$	Trade-off size & quality with a "slightly" dominant alternative	A	B	B
s12	$\{(80,80);(69,91);(71,89)\}$	$\{(81,92)\}$	Trade-off size & quality with a dominant alternative	A	B	B
s13	$\{(70,70);(50,71);(72,65)\}$	$\{(81,92)\}$	Trade-off size & quality with a "very" dominant alternative	A	B	B
Non-singleton sets of same size						
s14	$\{(80,75);(80,80)\}$	$\{(80,75);(70,90)\}$	Non-singleton sets of same size without a dominant alternative	∼	A/B/∼	∼
s15	$\{(80,75);(80,80);(81,91)\}$	$\{(80,75);(70,90);(81,91)\}$	Non-singleton sets of same size with a dominant alternative	∼	∼	∼

Notes: A pair (x_1,x_2) describes a hospital where x_1 corresponds to staff quality (% rating) and x_2 to comfort quality (% rating). Indirect utility's theoretical response pattern respects dominance relations; for s1, s4 to s6, and s14, IU ranking takes only one of the shown options depending on participants' response to which hospital they prefer (Q2).

Table I. Opportunity set comparisons and theoretical response patterns for cardinality, IU, and range of opportunity rules

III.E Procedures

Results are based on data from 4,902 participants. The mean and median time to complete the survey are respectively 912 ($\approx$ 15 minutes) and 666 seconds ($\approx$ 11 minutes). All these participants finished the survey, responded correctly to an attention question, and passed a speeding check (i.e., took more than 4 minutes and 15 seconds to complete the survey). Otherwise, participants were excluded from the data.[15]

We ran the study in ten countries: Brazil (BR), China (CH), Colombia (CO), France (FR), Japan (JA), Netherlands (NL), Portugal (PT), Turkey (TR), United Kingdom (UK), and the USA (US). We collected data in March 2021 using the survey company *Odities Technologie*, which sent an invitation email to its panel of participants to answer our survey. For completing the survey, participants received "tokens" that they could exchange for money. To minimize language effects, instructions in English were translated into the local language by professional native speakers, and back-translated to English by another person. Translators were careful to write the instructions in neutral language. The sample is representative of each country in terms of age and gender. Sample characteristics for each country are shown in Appendix C. There is heterogeneity between countries among most observable characteristics, which we control for in our cross-country analysis.

These countries were chosen for two reasons. First, for implementation purposes. In particular, at least one author is fluent in the language of 8 out of 10 of these countries. With the help of two additional colleagues, this allowed us to ensure the quality of the translations for all countries. Second, these countries were chosen for their diversity in terms of social, cultural, and institutional backgrounds. For instance, these countries differ in terms of their dominant religions and political institutions, which may translate into attitudinal differences. The attitudinal differences across these countries are illustrated by

[15]The attention question was randomly presented in the sequence of set comparisons. It consisted of a similar screen to the other set comparisons with an answer option "If you are not a robot please click on this button" (see instructions in Appendix G). The check of 4 minutes and 15 seconds was agreed upon with the survey company based on a pilot. Note that since our main research design does not rely on multiple treatments, differential attrition is not an issue. Still, potential differences in attrition across countries is one of the reasons why we control for observed characteristics in our cross-country analysis (see Appendix C for attrition data).

the fact that they are spread all over the influential Inglehart-Weltzel world cultural map (see World Values Survey 2022). We therefore believe that this selection of countries provides a somewhat comprehensive (even if clearly incomplete) test for potential cross-country differences.

IV. Empirical analysis and results

We structure our analysis as follows. First, we look at how people rank opportunity sets (Section IV.A). We show both aggregate response patterns and our main (Bayes) classifications for freedom and welfare. Second, we investigate if participants attach intrinsic value to freedom of choice (Section IV.B). For these two sections, we use the data from all 4,902 subjects. In Section IV.C, we test for cross-country differences.

IV.A How people rank opportunity sets

IV.A.1 Aggregate response patterns

We start this part of the analysis by showing the aggregate response patterns for the Freedom and Welfare questions. These are summarized in Table II. Some relevant patterns are already apparent from this table. On the one hand, there are questions for which there is considerable agreement, such as FoC in s4, s6, and s12 as well as welfare in s2, s3, and s13. On the other hand, there is considerable disagreement in others, such as FoC in s2 and s3 as well as welfare in s1, s7-s9, and s14. In addition, it is noticeable that responses to the Freedom and Welfare questions differ significantly. Note that these differences are in the expected direction. For example, most (though not all) people consider that *A* offers more FoC than *B* in s11 to s13, while most (though far from all) people consider that *B* offers more overall welfare than *A* in these set comparisons.[16]

[16] The patterns in Table II strongly suggest that answers are not random. Note that answers across sets for the same question (either Freedom or Welfare) also support non-randomness. For example, as expected, it is more often the case in s7 than in s8 and in s8 than in s9 that set *B* is considered to have more FoC than set *A*. Noise — which is common in similar studies — also seems moderate in our setting. For example, in s3 there are only 2.5% (10.6%) that state that *B* provides more (same) welfare than *A* when clearly set *A* dominates set *B*.

	Set comparisons		Freedom question: Which individual do you think has more freedom of choice?			Welfare question: Which individual do you think has the best insurance plan?		
	A	B	A	B	Same	A	B	Same
			(% all answers)					
Singleton sets								
s1	{(80,80)}	{(70,90)}	24.2	10.5	65.3	47.1	14.4	38.5
s2	{(81,92)}	{(80,75)}	40.0	3.1	56.9	81.7	3.0	15.3
s3	{(80,80)}	{(60,60)}	42.2	2.7	55.0	86.9	2.5	10.6
Adding an option								
s4	{(80,75)}	{(80,75),(71,89)}	6.0	78.6	15.4	7.8	60.0	32.2
s5	{(69,91),(71,89)}	{(69,91),(71,89),(70,90)}	6.1	72.9	21.0	8.1	52.2	39.8
s6	{(70,90),(71,89)}	{(70,90),(71,89),(80,75)}	7.0	75.3	17.7	9.8	59.5	30.8
s7	{(70,90),(71,89)}	{(70,90),(71,89),(68,88)}	8.6	66.9	24.4	14.5	40.7	44.8
s8	{(70,90),(71,89)}	{(70,90),(71,89),(60,60)}	10.4	63.9	25.6	18.8	35.6	45.7
s9	{(70,90),(71,89)}	{(70,90),(71,89),(20,30)}	14.9	55.2	29.9	27.6	24.6	47.7
s10	{(80,75),(95,95)}	{(80,75),(95,95),(71,89)}	8.4	67.6	24.0	10.5	25.2	64.3
Trade-off between size and quality								
s11	{(79,90),(77,91),(80,87)}	{(81,92)}	72.8	17.0	10.2	28.7	51.8	19.5
s12	{(80,80),(69,91),(71,89)}	{(81,92)}	73.0	18.0	8.9	27.7	55.4	16.9
s13	{(70,70),(50,71),(72,65)}	{(81,92)}	68.5	24.0	7.5	15.9	72.5	11.6
Non-singleton sets of same size								
s14	{(80,75),(80,80)}	{(80,75),(70,90)}	21.6	15.8	62.6	38.6	17.9	43.5
s15	{(80,75),(80,80),(81,91)}	{(80,75),(70,90),(81,91)}	15.4	13.1	71.5	22.4	16.2	61.4

Table II. Subjects' actual response patterns for Freedom and Welfare questions

IV.A.2 Main classifications

Each ranking rule offers a prediction in terms of which set provides more freedom/welfare for each set comparison. The *theoretical response patterns* of the most prominent rule of each family are summarized in Table I (see Table A.2 in Appendix A for the theoretical response patterns of all rules).

We can then compare the rules' theoretical response patterns to the subjects' actual response patterns in order to "classify" participants according to the rule they implicitly employ. We do this for both questions (Freedom and Welfare). Here we follow Ambuehl and Bernheim (2021) who apply the methods from Hastie et al. (2001) and Costa-Gomes et al. (2001) to perform a Bayes classification exercise similar to ours. We assign to each subject a rule among the 23 plausible theoretical rules reviewed above and an error probability. To do that, we use the posterior probability to follow a rule with a certain error probability conditional on subjects' response patterns. Formally, denote by c_i the response pattern of subject i, i.e., the vector of answers that subject i gives to the $S = 15$ binary set comparisons. Denote by $P(R_j, \varepsilon_j | c_i)$ the posterior probability to follow a rule R_j with error probability ε_j conditional on the response vector c_i, with $j \in \{1, ..., J\}$ and $J = 23$ in our analysis. We assign a subject to rule R_j if and only if:

$$P(R_j, \varepsilon_j^* | c_i) > P(R_k, \varepsilon_k^* | c_i) \quad \forall \quad k \in \{1, ..., J\} \setminus j \tag{1}$$

where ε_j^* is the probability that maximizes $P(R_j, \varepsilon_j | c_i)$, i.e., ε_j^* is the probability that maximizes the likelihood that a subject i with response vector c_i follows the rule R_j. When more than one rule maximizes the posterior probability, we assign the subject to a rule at random. See Appendix B for further technical details.

Classification results for Freedom

The main classification results for the Freedom question are displayed in Figure I. The cardinality rule is the most often implicitly used by subjects: 38% of subjects make choices that are consistent with this rule. In addition, the weak cardinality rule gathers 8% of subjects, and 20% have a response pattern consistent with a lexicographic rule that gives priority to the menu size. All in all, a large majority (68%) of participants rank sets in terms of FoC according to cardinality family rules. On the contrary, quality-based rules are seldom used. The IU family accounts for 11% of participants, while the potential preferences family accounts for only 3% of subjects.

How good is the fit between our classification and the subjects' choices? To see this, we look at the average error (ε^*) for each ranking rule, i.e., the average proportion of subjects' responses that contradict the rule assigned to them by the Bayesian classifier. The first row of Table III reports the average error for the main rules and families for the Freedom question. The total average error is 0.21 (i.e., on average subjects choose differently than the rule assigned to them in around 3 out of 15 set comparisons). Note that this error is considerably lower for the most prevalent rule: subjects assigned to the cardinality rule only contradict it in 1.5 set comparisons on average. Overall, the fit is very good for the 46% who follow cardinality and weak cardinality, good for a significant proportion of other subjects (such as the 24% assigned to Lex Cardinality-RoO, Lex Cardinality-IU, and IU), and less good for the least prevalent rules.

Table IV shows the choices for the different set comparisons for participants that follow the cardinality rule. This data allows us to identify where subjects depart from this rule. Three insights come out of it. First, the goodness-of-fit is high for all questions, and for 12 out of 15 questions the consistency is larger than 92%. Second, not surprisingly, it seems that participants who follow cardinality are most likely to depart from it when we add an option of "poor" quality (set comparison s9). Third, deviations are also more common when these subjects compare non-singleton sets with the same size (see s14 and s15). While some of these deviations may be due to random mistakes, this suggests that a small number of subjects assigned to cardinality are sensitive to quality considerations in these three set comparisons.

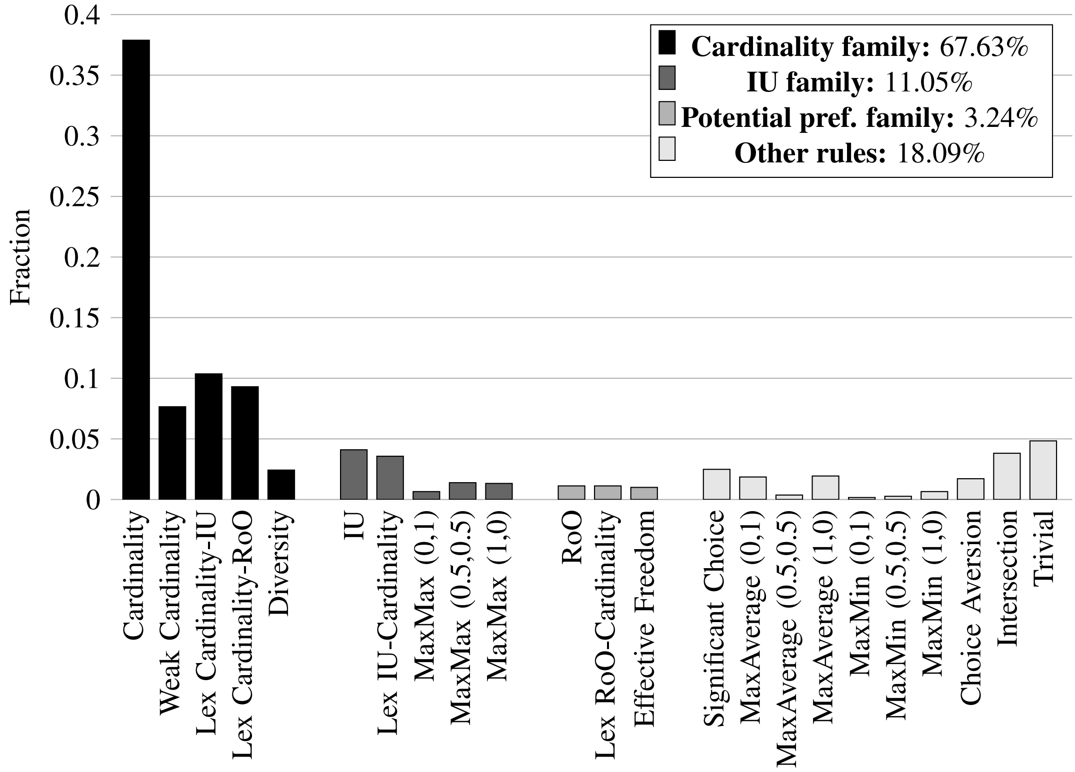


Figure I. Bayesian classification for Freedom question

Classification results for Welfare

The main classification results for the Welfare question are displayed in Figure II. As can be seen from the figure, the results are much less clear than for the Freedom question. The most prevalent rule is IU. However, only 14% of subjects respond as if following this rule. Despite the heterogeneity of the classification, two patterns are apparent. First, a significant number of subjects follow quality-based rules. In particular, 37% of participants follow IU family rules (14% IU, 12% Lex IU-Cardinality, and 12% MaxMax) and 16% are consistent with potential preferences family rules. Second, cardinality-based rules are seldom used to evaluate the welfare of sets, with only 16% of subjects following a cardinality family rule for this question (with most of these assigned to Lex Cardinality-IU and Lex Cardinality-RoO).

It is also worth noting that these classification results have less goodness-of-fit than the ones for the Freedom question. As reported in the second column of Table III, the average error is 0.30 for the Welfare question. This error is lower for the most prevalent

Question	**Average**	**Cardinality family**	Cardinality	Weak cardinality	Lex Cardinality-IU	Lex Cardinality-RoO	Diversity	**IU family**	Indirect Utility	Lex IU-Cardinality	MaxMax (0,1)	MaxMax (0.5,0.5)	MaxMax (1,0)	**Potential Pref. family**	**Other rules**
Freedom	0.21	0.15	0.10	0.08	0.31	0.24	0.29	0.33	0.23	0.38	0.43	0.39	0.39	0.36	0.34
Welfare	0.30	0.33	0.26	0.23	0.35	0.33	0.65	0.28	0.21	0.32	0.33	0.28	0.32	0.27	0.33

Notes: This table reports the average error (ε^*) for main ranking rules and families, i.e., the average number of subjects' responses that contradict the rule assigned to them by the Bayesian classifier. "Cardinality family", "IU family", and "Potential Pref. family" report the average error for the rules in the respective families, while "Other rules" reports the average error for the theoretical rules that do not belong to these families.

Table III. Goodness-of-Fit of the Bayesian classifier

Comparisons:	s1	s2	s3	s4	s5	s6	s7	s8	s9	s10	s11	s12	s13	s14	s15
Prediction:	∼	∼	∼	B	B	B	B	B	B	B	A	A	A	∼	∼
Choice freq.:															
A	5	5	5	2	1	2	2	3	4	2	98	98	96	7	7
∼	92	94	94	1	0	3	3	4	11	3	1	1	1	85	87
B	3	1	1	97	99	95	95	93	85	95	1	1	3	8	6

Table IV. Choices made by subjects classified as Cardinality (Freedom question)

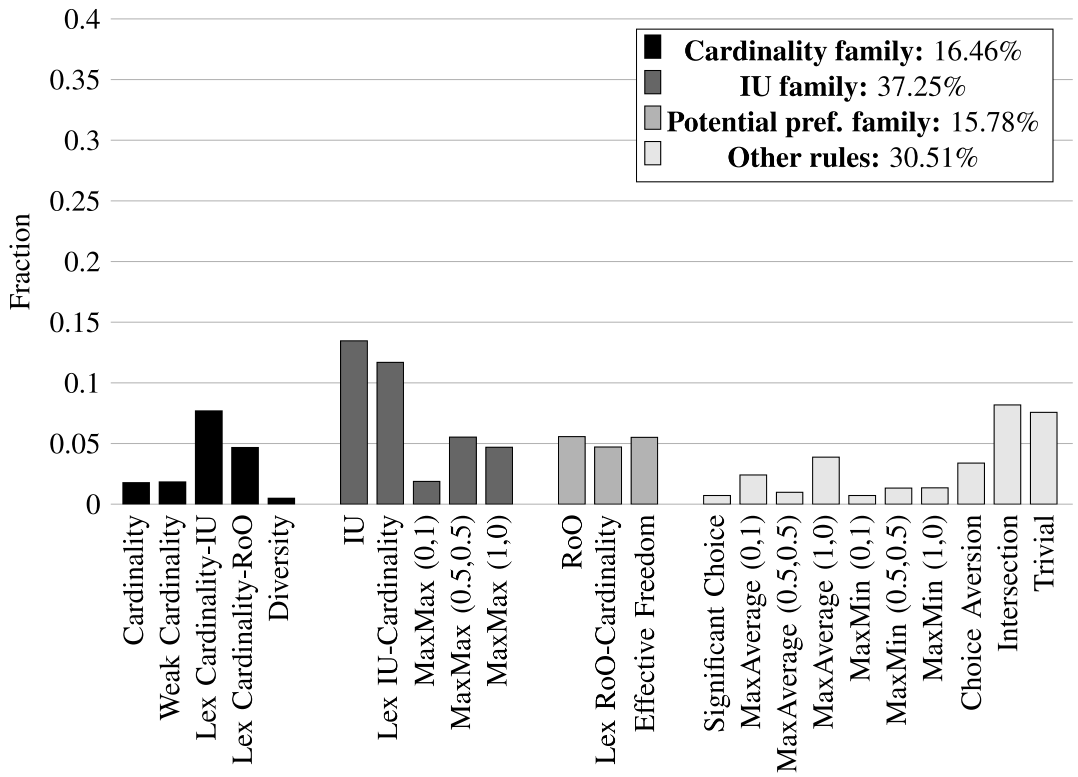


Figure II. Bayesian classification for Welfare question rule (0.21 for IU), and moderate for other quality-based rules belonging to the IU family (0.28 on average) and the potential preferences family (0.27 on average). Overall, it is apparent that subjects are more idiosyncratic in their welfare evaluations than in their freedom evaluations.

Finally, Table V reports the consistency between IU's rule predictions and the choices made by subjects classified as IU.[17] We can see that the goodness-of-fit is high for most questions. For the questions for which the consistency is lower, the most interesting pattern is from s9, where subjects assigned to IU who contradict it tend to prefer the smaller set (131 out of 158). This suggests that these subjects consider that a "very bad" alternative is detrimental to the quality of the set. This trend is aligned with our aggregate result that 28% of our sample prefers the smaller set in s9 (see Table II).

Comparisons:	s1	s2	s3	s4	s5	s6	s7	s8	s9	s10	s11	s12	s13	s14	s15
% of choices:	91	93	95	83	85	86	84	82	81	88	78	81	95	87	74

Table V. % of choices made by participants classified as IU consistent with IU

Robustness checks and alternative explanations

Several checks reported in Appendix D provide further confidence to our main classifications. First, our results are robust when we perform checks for attention. The classifications are very similar when we remove subjects that make clear mistakes in their comparisons of sets, such as stating that (60,60) is better than (80,80) (Appendix D.1).

Second, results are also very similar when we perform our analysis per quartiles of total response times (Appendix D.2). The latter analysis also shows that cardinality is *not* being used as a heuristic to quickly respond to the questions: contrary to this hypothesis, the fastest 25% of participants are significantly less likely to follow cardinality in the Freedom question. In the same direction, subjects assigned to the cardinality family are statistically significantly slower than others.[18]

[17] We cannot report a table equivalent to Table IV because the IU's rule predictions depend on subjects' preferences over hospitals (Q2).

[18] Average and median times to complete the survey are respectively 944 and 700 seconds for subjects

Third, as reported in Appendix D.3, our results are robust to the order of the questions (versions *FreedomIU* and *IUFreedom*). Consistent with order effects and the relevance of salience, we find that subjects in the freedom classification are statistically significantly less likely to be assigned to the cardinality rule in *IUFreedom* than *FreedomIU* (i.e., when they are first asked the IU question). However, this effect is small and does not change our main results (42% and 34% are assigned to cardinality in *IUFreedom* and *FreedomIU* respectively). In addition, the overall classifications of both freedom and welfare are otherwise very similar across the two versions.

Fourth, our classifications using subjects' responses do much better than a classification using random answers (Appendix D.4). In particular, with random answers, no rule gathers more than 9% of (artificial) subjects and the average error is 0.70 (compared to 0.21 and 0.30 for our freedom and welfare classifications respectively).

Fifth, a cluster analysis using a k-modes procedure to infer rules from the data leads to very similar results to the ones using the Bayesian classifier that fits rules to the data (see Chaturvedi et al. 2001; see Appendix D.5 for results). In particular, all our main findings hold. This analysis also suggests that, in line with results reported above, a few subjects exhibit choice aversion for particular set comparisons (namely for s7 to s9). In addition, it suggests that, when evaluating the welfare of sets, about 15% of subjects implicitly follow a version of IU for which one set is overall better than another when it contains a hospital that clearly dominates the best hospital of the other set (as in s2, s3, and s13), otherwise both sets provide the same overall welfare.

Finally, our main results hold when we restrict the analysis to set comparisons where it is easy to infer Mr. Green and Mr. Yellow's preferences over alternatives (s2, s3, s7 to s13 and s15; see Appendix D.6 for results). This shows that our results are not driven by the uncertainty about others' preferences.

---

assigned to the cardinality family and 845 and 586 seconds for others ($p = 0.0085$ for a Student test for difference between means and $p < 0.001$ for a Wilcoxon-Mann-Whitney test).

	Relevant cases		IvFoC ($C \succ_i^W D$ in relevant cases)					
				% of subjects s.t. IvFoC revelations				
	Mean (out of 15)	% subjects >0	Mean	$=0$	≥ 1	≥ 2	≥ 3	≥ 4
$C \succ_i^F D$ and								
$\max(D) R_i \max(C)$	7.15	96.43	3.42	15.12	84.88	70.60	57.67	45.84
$\max(D) I_i \max(C)$	5.13	95.19	2.87	17.50	82.50	66.67	51.96	38.43
$\max(D) P_i \max(C)$	2.02	80.74	0.54	66.10	33.90	14.46	4.92	0.78

Notes: The first column shows the mean number of responses per subject (in a total of 15 set comparisons) such that C (either set A or B) is considered to have more freedom than D (A if $C = B$ or B if $C = A$) and $\max(D)R_i \max(C)$ (first row), $\max(D)I_i \max(C)$ (second row), or $\max(D)P_i \max(C)$ (third row). The second column shows the percentage of participants for which the number of relevant cases is at least one. The third column shows the mean number of responses per subject such that $C \succ_i^W D$ in the relevant cases (i.e., when $C \succ_i^F D$ and $\max(D)R_i \max(C)$/$\max(D)I_i \max(C)$/$\max(D)P_i \max(C)$). The last five columns present the percentage of subjects who reveal IvFoC among the participants for which the number of relevant cases is at least one.

Table VI. IvFoC: Main results

IV.B The intrinsic value of freedom

Table VI reports our main results for the intrinsic value of freedom. The first two columns suggest that there is a significant number of "relevant cases", i.e., set comparisons for which subjects express that there is a "conflict" between FoC and the "best" alternative. In a total of 15 set comparisons, the average number of situations with a "conflict" between FoC and the best alternative is close to 50% (7.15 out of 15) and almost all subjects (96%) are presented with at least one set comparison for which they answered that $C \succ_i^F D$ and $\max(D)R_i \max(C)$. From these, the majority are cases in which subjects are indifferent between the best element of C and D [$\max(D)I_i \max(C)$], but there is a significant number of cases where subjects strictly prefer the best element of D to that of C [$\max(D)P_i \max(C)$]. The latter cases present a clear conflict between FoC and the best alternative, while the former present a "mild" conflict in the sense that a subject can still contradict standard economic theory by revealing that he/she attaches intrinsic value to freedom.

How many subjects reveal attaching intrinsic value to freedom *in these relevant cases*? The results are striking. On average, 48% of answers reveal giving intrinsic value to freedom in these situations (3.42 out of 7.15). Moreover, 85% of subjects reveal that they attach intrinsic value to freedom at least once and a majority of subjects (58%) reveal it at least three times. A large proportion of these IvFoC revelations are for cases in which participants are indifferent between the best alternative of the two sets (second row of Table VI). Still, there is a significant number of subjects that express that a larger set provides more overall welfare than a smaller set even though the larger set has a "worse" (dominated) top alternative than the smaller set. In fact, 34% of subjects reveal attaching such "high" intrinsic value to freedom at least once when, on average, they could only show it at most twice (see third row of the first and fifth columns of Table VI). Overall, these results contradict the standard view according to which freedom has only instrumental value.

Table VII reports the proportion of participants who reveal attaching IvFoC for each set comparison. This can bring further insights into when and why people exhibit this attitude. Results can be summarized as follows. First, it is difficult to rationalize the answers that attach IvFoC when comparing singleton sets (s1 to s3). However, as shown in the table, there are very few subjects that do so (41.37%*278=115 for s1, 69 for s2, and 53 for s3). Second, the table shows that for set comparisons in which we "add an option" (s4 to s10), the IvFoC is driven by cases in which participants are indifferent between the best alternative of the two sets (i.e., s4 to s10 in columns ii). This is not surprising. For instance, in s7 to s10 the "best" alternative is present in both sets. Overall, IvFoC ranges from 33% to 69% of relevant cases on these set comparisons (s10 and s4 respectively in columns ii), and its prevalence seems to depend on the quality of the top alternative *and* the alternative being "added" to the set. Third, set comparisons with a "trade-off between size and quality" (s11 to s13) are behind the above-stated result that there is a significant proportion of participants that give so much intrinsic value to freedom that they prefer a larger set even though it has a "worse" top alternative (see s11 to s13 in columns iii). Here again, the quality of the alternatives seems to matter. Finally, fewer participants find a "conflict" between freedom and the best alternative when comparing non-singletons of the same size (s14 and s15). Nonetheless, even for these cases, there is a non-negligible number of participants who seem to judge the overall welfare of the sets

Set comparisons			IvFoC ($C \succ_i^W D$ in relevant cases)					
			(i) $C \succ_i^F D$ and $\max(D) R_i \max(C)$		(ii) $C \succ_i^F D$ and $\max(D) I_i \max(C)$		(iii) $C \succ_i^F D$ and $\max(D) P_i \max(C)$	
	A	B	% IvFoC	# relevant cases	% IvFoC	# relevant cases	% IvFoC	# relevant cases
Singleton sets								
s1	{(80,80)}	{(70,90)}	41.37	278	60.47	43	37.87	235
s2	{(81,92)}	{(80,75)}	48.59	142	60.00	35	44.86	107
s3	{(80,80)}	{(60,60)}	41.73	127	66.67	21	36.79	106
Adding an option								
s4	{(80,75)}	{(80,75),(71,89)}	69.06	3,038	69.23	2,896	65.49	142
s5	{(69,91),(71,89)}	{(69,91),(71,89),(70,90)}	62.51	2,942	62.89	2,881	44.26	61
s6	{(70,90),(71,89)}	{(70,90),(71,89), (80,75)}	69.04	2,571	68.69	2,427	75.00	144
s7	{(70,90),(71,89)}	{(70,90),(71,89),(68,88)}	58.66	3,643	58.75	3,634	22.22	9
s8	{(70,90),(71,89)}	{(70,90),(71,89),(60,60)}	53.47	3,585	53.48	3,573	50.00	12
s9	{(70,90),(71,89)}	{(70,90),(71,89),(20,30)}	48.78	3,403	48.78	3,393	50.00	10
s10	{(80,75),(95,95)}	{(80,75),(95,95),(71,89)}	34.26	3,657	33.44	3,595	82.26	62
Trade-off between size and quality								
s11	{(79,90),(77,91),(80,87)}	{(81,92)}	30.88	3,177	48.02	429	28.20	2,748
s12	{(80,80),(69,91),(71,89)}	{(81,92)}	29.02	3,218	50.00	370	26.30	2,848
s13	{(70,70),(50,71),(72,65)}	{(81,92)}	18.18	3,207	49.73	185	16.25	3,022
Non-singleton sets of same size								
s14	{(80,75),(80,80)}	{(80,75),(70,90)}	60.33	842	71.81	518	41.98	324
s15	{(80,75),(80,80),(81,91)}	{(80,75),(70,90),(81,91)}	67.85	1,238	68.46	1,151	59.77	87
All			47.80	35,068	56.03	25,151	26.93	9,917

Notes: This table reports the % of participants that prefer C (either set A or B) to D (A if $C = B$ or B if $C = A$) given that they consider that C provides more freedom than D while the preferred element(s) of D is (i) weakly preferred, (ii) indifferent to, or (iii) strictly preferred to that of C. "# relevant cases" denotes the number of responses/participants per set comparison such that $C \succ_i^F D$ and $\max(D) R_i \max(C)$/$\max(D) I_i \max(C)$/$\max(D) P_i \max(C)$.

Table VII. IvFoC: Results per set comparison taking into account more information than its top alternative.

Robustness checks and alternative explanations

Several checks provide further support to these findings (see Appendix E). First, as with our classification results, these findings are very similar when we remove subjects that made clear mistakes in their comparisons of sets (Appendix E.1), and when we perform our analysis per quartiles of total response times (Appendix E.2). Second, the IvFoC is robust to the order of the questions (versions *FreedomIU* and *IUFreedom*). Once again, even though there are statistically significant order effects, the overall results are strikingly similar (Appendix E.3).

Our results are also robust to two potential alternative explanations. First, *preferences for flexibility* (i.e., a preference for larger sets to better tailor for multiple potential future preferences), could, at least in principle, explain the intrinsic value of freedom in some of our set comparisons. This could be problematic, since preferences for flexibility are an instrumental concern about the benefits that FoC can entail in the future. However, in our setting, preferences for flexibility could only explain IvFoC in set comparisons where a non-dominated alternative is added to the set (s4 to s6). They cannot rationalize the behavior of a participant who reveals IvFoC in set comparisons where dominated alternatives are added (s7 to s10). They can neither explain revelations of IvFoC in other set comparisons such as s11 to s13 and s15. However, Table VII shows that IvFoC is revealed in 40% of the relevant cases in the latter set comparisons (s7 to s13 and 15). It follows that preferences for flexibility cannot be the underlying reason for the intrinsic value of freedom in our setting.[19]

Second, participants could potentially compare opportunity sets based on hospital attributes that are not mentioned in our vignette. For example, they could think that a larger number of hospitals increases the likelihood of having hospitals close to one's address.

[19]Differences in levels between s4 to s6 and s7 to s10 are not directly comparable since the alternatives added are not of similar quality. We therefore abstain from making an inference if preferences for flexibility can or cannot explain a small percentage of observed IvFoC.

If this would be the case, a participant could reveal attaching IvFoC (in the above sense) for instrumental reasons, which would go against our interpretation. Take the example of s4, with menu A with one hospital (80,75) versus menu B with two hospitals (80,75) and (71,89). Assume the subject prefers (80,75) to (71,89), thinks menu B offers more freedom than menu A, and considers that menu B provides more overall welfare than A. This is a revelation of IvFoC in the above sense. This pattern of responses would be consistent with the alternative explanation based on proximity *only if* the subject would go to a closer hospital with (71,89) than a further hospital with (80,75), even though (80,75) is preferred to (71,89). While this alternative explanation could rationalize IvFoC in some set comparisons, it is very unlikely to explain it in some others. For instance, it is very unlikely that having the hospital $(20, 30)$ in s9 is valued for instrumental (proximity) reasons. Since we observe almost 50% of IvFoC revelations in the relevant cases of the latter set comparison, this suggests that something intrinsic about having more choice is driving most of the observed intrinsic value of freedom.

IV.C Cross-country results

Are these results dependent on the social, cultural, and institutional background of participants? To answer this question, we compare our main results across 10 countries with very different social, cultural, and institutional backgrounds. Since our subject pools may differ across countries due to differences in recruitment and other socio-economic characteristics, we control for observed characteristics in our comparison across countries. In accordance, we present results for the "average participant" (i.e., a participant whose characteristics are averaged over all countries).

IV.C.1 Main classifications

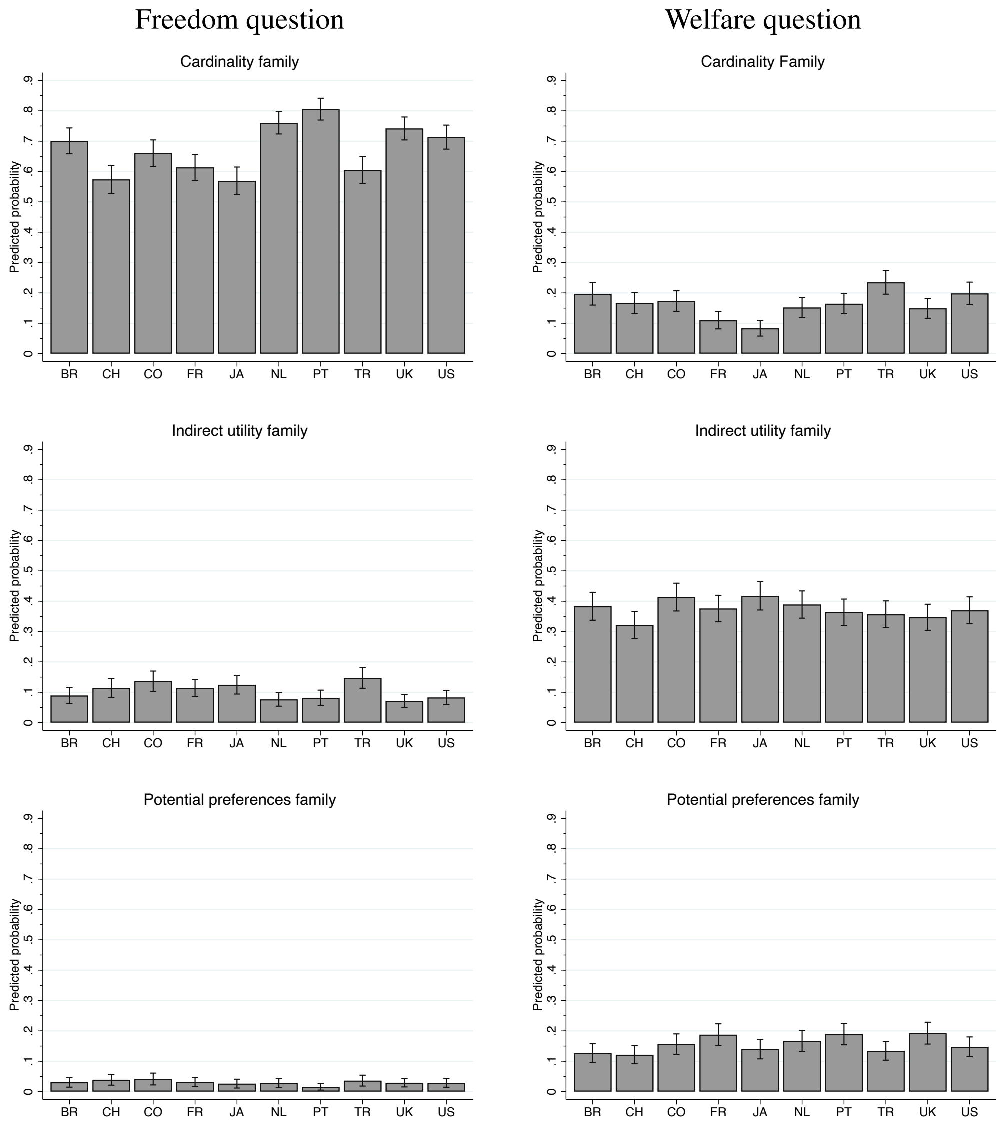
To test if people from different countries rank sets differently, we first estimate a multinomial logit regression with the assigned rule as the dependent variable. We use countries and observed characteristics as independent variables. We then estimate, for each country, the predicted probability of the average participant to be assigned to a given rule as if he/she is from that country.

Figure III reports, for each country, the predicted probability of an average participant to be assigned to a rule from the cardinality family, IU family, or potential preferences family. For the Freedom question (left panel), between 57% (Japan and China) and 81% (Portugal) of participants per country belong to the cardinality family. Some of the country differences are statistically significant (see Table F.2 in Appendix F). On the other hand, the IU and potential preferences families account together for less than 20% in all countries. Therefore, despite some visible differences, the overall message is consistent across countries: A majority of subjects rank the freedom of choice in menus as if following size-based rules, while only a small minority follows quality-based rules when evaluating the freedom of sets.

For the Welfare question (right panel), the IU family is the most prevalent across countries (between 32% in China and 42% in Japan), while the potential preferences family gathers between 12% (China) and 19% (UK) of subjects across countries. Most of these differences are small and most (though not all) are statistically insignificant at 5% (see Table F.3 in Appendix F). On the other hand, there are between 8% (Japan) and 24% (Turkey) participants across countries that rank the welfare of sets following size-based rules. Again, despite some visible differences, the message is robust across countries: A significant proportion of participants use IU-based rules and other quality-based rules to judge the welfare of sets, but there is considerable heterogeneity in how people evaluate the welfare of sets.

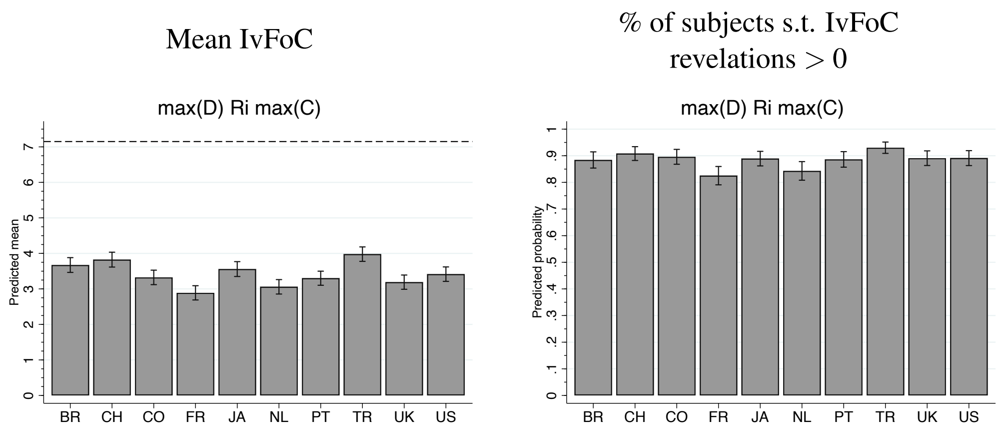
IV.C.2 Intrinsic value of freedom

Figure IV reports, for each country, the predicted average number of times that the average participant reveals IvFoC and the predicted probability that an average participant exhibits IvFoC at least once. As shown in the figure, the overall picture is very similar across countries. In all countries, the predicted average number of times that the average



Notes: These figures report the predicted probabilities of an average participant to be assigned to the cardinality family, IU family, and potential preferences family for the Freedom question (left panel) and the Welfare question (right panel). These estimates are based on multinomial logit regressions controlling for observed characteristics (see Table F.1 in Appendix F). Lines indicate 95% confidence interval.

Figure III. Classifications per country



Notes: The left panel reports the predicted average value of the number of responses such that C (either set A or B) is considered to provide more welfare than D (A if $C = B$ or B if $C = A$), when C is considered to have more freedom than D and $\max(D)R_i\max(C)$, using an OLS regression controlling for the number of relevant cases and observed characteristics. The right panel reports the predicted probability of this number being positive, using an OLS regression controlling for the same variables. See Table F.6 in Appendix F for underlying regressions. Lines indicate 95% confidence interval.

Figure IV. Intrinsic value of freedom per country

participant reveals IvFoC is between 2.89 (France) and 3.98 (Turkey) (left panel). Although many of these differences are statistically significant (see Table F.4 in Appendix F), the magnitude of these differences is rather small. Results are again very similar when we look at the predicted probability that an average participant reveals IvFoC at least once (right panel). Even though some of these differences are statistically significant, the effect size does not question our main findings (see Table F.5 in Appendix F).[20] Overall, these results demonstrate that attaching intrinsic value to freedom of choice in our setting is a robust cross-cultural phenomenon independent of participants' social, cultural, or institutional background.

V. CONCLUDING REMARKS

It is standard practice in economics to evaluate social states based on individual preferences. "Freedom" is often mentioned as an important value, but only for instrumental reasons: giving individuals freedom of choice makes it possible for them to pick the outcomes they consider the best in terms of their individual preferences, which are largely unknown to the policymaker. This almost exclusive focus on outcomes, evaluated in

[20]In Appendix F, we show that differences across countries are mainly driven by cases where participants are indifferent between the top alternatives of both sets.

terms of preferences, seems at odds with public debates in several countries where "freedom" is frequently mentioned and valued without reference to preferences.

Our paper sheds new light on this topic. We use a novel survey-based research design to i) understand if people attach intrinsic value to freedom of choice, and ii) check which theoretical rules people implicitly employ when ranking opportunity sets in terms of freedom and overall welfare. We do this for a total of 4902 participants across 10 countries with distinct social, cultural, and institutional backgrounds. Surprisingly, a majority of subjects reveal that they attach intrinsic value to freedom. We also find that a large majority of subjects use size-based rules to rank sets in terms of freedom, while there is considerable heterogeneity in the rules that subjects implicitly employ to rank sets in terms of welfare. These results are strikingly robust across countries.

Results in positive welfare economics can help find a balance between people's intuitions and theoretical principles (see e.g. Gaertner and Schokkaert 2012; Ambuehl and Bernheim 2021). Our results provide new and important insights in this direction. On the one hand, they provide support for some theoretical viewpoints. For example, our results support the view that counting the number of alternatives is "a natural way of measuring freedom in the absence of information about the agent's preferences" (Foster 2011, p. 20). On the other hand, our results question some other viewpoints. For instance, people's potential preferences and the "eligibility" of options — two notions widely supported in the FoC literature — do not seem to drive people's views about which options increase freedom of choice. Similarly, contrary to the outcome-based dominant view, our results show that welfare evaluations of opportunity sets are multidimensional and differ significantly across people. This latter insight suggests that while evaluating sets by their indirect utility may be a sensible and parsimonious simplification in some settings, that approach will disregard relevant welfare-enhancing features of opportunity sets and considerable heterogeneity in the population (see Benjamin et al. 2014 for related evidence). Taken together, these findings provide valuable input to new positive and normative theories of how people rank (or should rank) opportunity sets.

From a policy perspective, many authors have argued that public policy should *not* be based on satisfying preferences, but instead on providing opportunities for individuals to achieve their own ends, whatever those ends might be (e.g. Rawls 1971; Roemer 1998; Sugden 2004). Even if one does not endorse this view, people's preferences over menus can still *complement* information about preferences over alternatives as inputs to public policy. In both cases, our results can inform policymakers on how to compare different opportunity sets if they wish to respect people's preferences over menus. For example, our results point out that people value freedom of choice for its own sake. In particular, we show that it is important to enlarge small opportunity sets even if this does not lead to different choices. This finding supports policy reforms that provide greater choice in small sets, as these reforms can be beneficial even for people who will not change their behavior. An example is found in the National Health Service in the UK, which has changed over the years how many hospitals — from one to five — patients can receive treatment in (most recent reform by the UK Government in 2023). An important goal for subsequent research is to understand to what extent the intrinsic value of freedom holds for stakeholder decisions, larger sets, and different contexts.

João V. Ferreira, University of Southampton, UK.

Nobuyuki Hanaki, Institute of Social and Economic Research, Osaka University, Japan.

Fabrice Le Lec, Univ. Lille, CNRS, IESEG School of Management, UMR 9221 LEM Lille Économie Management, France.

Erik Schokkaert, Department of Economics, KU Leuven, Belgium.

Benoît Tarroux, University Lumière Lyon 2 and GATE Lyon-Saint-Etienne, France.

ACKNOWLEDGMENTS

We are grateful to Sandro Ambuehl, Alexander Cappelen, Olivier L'Haridon, Yannick L'Horty, Stephane Luchini, Séverine Toussaert, Michael Vlassopoulos, and participants of workshops in Lyon, Southampton, and Aussois (TEPP winter seminar), for valuable discussions, comments and suggestions. We are also grateful to professional and friend translators and Sizar Mani from Odity for data collection. This paper forms part of the research project "ValFree" (The Value of Choice, grant No. ANR-16-CE41-0002-01) of the French National Agency for Research whose financial support is gratefully acknowledged. Hanaki gratefully acknowledges financial support from grants-in-aid for scientific research (KAKENHI, grant numbers: 18K19954, 20H05631) from the Japan Society for the Promotion of Science (JSPS) as well as support from the Joint Usage/Research Center at ISER, Osaka University.

Declaration of interests

None to declare.

References

Ackfeld, V. and Ockenfels, A. (2021). Do people intervene to make others behave prosocially? *Games and Economic Behavior*, 128:58–72.

Almås, I., Cappelen, A., and Tungodden, B. (2020). Cutthroat capitalism versus cuddly socialism: Are americans more meritocratic and efficiency-seeking than scandinavians? *Journal of Political Economy*, 128:1753–88.

Alsan, M., Braghieri, L., Eichmeyer, S., Kim, J., Stantcheva, S., and Yang, D. (2023). Civil liberties in times of crisis. *AEJ: Applied Economics*, (forthcoming).

Ambuehl, S. and Bernheim, B. D. (2021). Interpreting the will of the people: a positive analysis of ordinal preference aggregation. *NBER Working Paper*, (29389).

Ambuehl, S., Bernheim, B. D., and Ockenfels, A. (2021). What motivates paternalism? an experimental study. *The American Economic Review*, 111(3):787–830.

Arad, A. and Rubinstein, A. (2018). The people's perspective on libertarian-paternalistic policies. *The Journal of Law and Economics*, 61(2):311–33.

Arlegi, R., Bourgeois-Gironde, S., and Hualde, M. (2022). Attitudes toward choice with incomplete preferences: An experimental study. *Journal of Economic Behavior & Organization*, 204:663–79.

Arrow, K. J. (1995). A note on freedom and flexibility. In Pattanaik, K. B. P. K. and Suzumura, K., editors, *Choice, Welfare, and Development, A Festschrift in Honour of Amartya K. Sen*. Oxford University Press, Oxford.

Bansak, K., Hainmueller, J., Hopkins, D. J., and Yamamoto, T. (2018). The number of choice tasks and survey satisficing in conjoint experiments. *Political Analysis*, 26(1):112–9.

Barbera, S., Bossert, W., and Pattanaik, P. K. (2004). Ranking sets of objects. In Barbera, S., Hammond, P., and Seidl, C., editors, *Handbook of Utility Theory*, pages 893–977. Springer.

Bartling, B., Fehr, E., and Herz, H. (2014). The intrinsic value of decision rights. *Econometrica*, 82(6):2005–39.

Basu, K. and López-Calva, L. (2011). Functionings and capabilities. In Arrow, K., Sen, A. K., and Suzumura, K., editors, *Handbook of Social Choice and Welfare*, volume II, pages 153–88. Elsevier.

Baujard, A. (2007). Conceptions of freedom and ranking opportunity sets – a typology. *Homo Oeconomicus*, 24:231–54.

Benjamin, D. J., Heffetz, O., Kimball, M. S., and Szembrot, N. (2014). Beyond happiness and satisfaction: Toward well-being indices based on stated preference. *The American Economic Review*, 104(9):2698–735.

Bervoets, S. and Gravel, N. (2007). Appraising diversity with an ordinal notion of similarity: an axiomatic approach. *Mathematical Social Sciences*, 53(3):259–73.

Bossert, W., Pattanaik, P. K., and Xu, Y. (1994). Ranking opportunity sets: an axiomatic approach. *Journal of Economic Theory*, 63(2):326–45.

Carter, I. (1999). *A Measure of Freedom, Oxford University Press.* Oxford University Press.

Chaturvedi, A., Green, P. E., and Caroll, J. D. (2001). K-modes clustering. *Journal of classification*, 18:35–55.

Costa-Gomes, M., Crawford, V. P., and Broseta, B. (2001). Cognition and behavior in normal-form games: An experimental study. *Econometrica*, 69(5):1193–235.

Costa-Gomes, M. A. and Crawford, V. P. (2006). Cognition and behavior in two-person guessing games: An experimental study. *American economic review*, 96(5):1737–1768.

Dowding, K. and van Hees, M. (2009). Freedom of choice. In Anand, P., Pattanaik, P. K., and Puppe, C., editors, *Handbook of Rational and Social Choice*, pages 374–92. Oxford University Press.

Dreyer, L. and Mahler, D. (2022). Free to choose or free to lose? understanding individual attitudes toward paternalism. *Behavioural Public Policy*, pages 1–23.

Ferreira, J. V., Hanaki, N., and Tarroux, B. (2020). On the roots of the intrinsic value of decision rights: Experimental evidence. *Games and Economic Behavior*, 119:110–122.

Foster, J. (2011). Freedom, opportunity and well-being. In Arrow, K., Sen, A. K., and Suzumura, K., editors, *Handbook of Social Choice and Welfare*, volume II, pages 687–728. Elsevier.

Gaertner, W. and Schokkaert, E. (2012). *Empirical Social Choice: Questionnaire-Experimental Studies on Distributive Justice*. Cambridge University Press, Cambridge.

Gravel, N. (2008a). What is diversity? In *Economics, rational choice and normative philosophy*. Routeledge Frontiers in Political Economy, London.

Gravel, N. (2008b). What is freedom? In *Handbook of Economics and Ethics*. Edward Edgar Publishing, London.

Gul, F. and Pesendorfer, W. (2001). Temptation and self-control. *Econometrica*, 69(6):1403–36.

Hastie, T., Tibshirani, R., and Friedman, J. (2001). *The Elements of Statistical Learning*, volume 1. Springer series in statistics, New York.

Hausman, D. M. (2012). *Preference, Value, Choice, and Welfare*. Cambridge University Press, New York.

Hayek, F. A. (1960). *The Constitution of Liberty*. Routledge., London.

Konow, J. (2009). Is fairness in the eye of the beholder? an impartial spectator analysis of justice. *Social Choice and Welfare*, 33:101–27.

Kreps, D. M. (1979). A representation theorem for “preference for flexibility”. *Econometrica*, 47:565–77.

Le Lec, F. and Tarroux, B. (2020). On attitudes to choice: some experimental evidence on choice aversion. *Journal of the European Economic Association*, 18(5):2108–34.

Mill, J. S. (1859). *On Liberty*. Penguin Classics, London, [2010] edition.

Müller, D. and Renes, S. (2021). Fairness views and political preferences: evidence from a large and heterogeneous sample. *Social Choice and Welfare*, 56(4):679–711.

Nehring, K. and Puppe, C. (2002). A theory of diversity. *Econometrica*, 70(3):1155–98.

Nozick, R. (1974). *Anarchy, State, and Utopia*. Basic Books.

Pattanaik, P. K. and Xu, Y. (1990). On ranking opportunity sets in terms of freedom of choice. *Recherches Economiques de Louvain*, 56:383–90.

Pattanaik, P. K. and Xu, Y. (1998). On preference and freedom. *Theory and Decision*, 44:173–98.

Pattanaik, P. K. and Xu, Y. (2000). On ranking opportunity sets in economic environments. *Journal of Economic Theory*, 93(1):48–71.

Puppe, C. (1996). An axiomatic approach to “preference for freedom of choice”. *Journal of Economic Theory*, 68:174–99.

Rawls, J. (1971). *A Theory of Justice*. Harvard University Press, Cambridge.

Robeyns, I. (2017). *Wellbeing, Freedom and Social Justice, the Capability Approach Re-Examined*. Open Book Publishers, Cambridge, UK.

Roemer, J. (1998). *Equality of opportunity*. Harvard University Press.

Sen, A. K. (1985). *Commodities and Capabilities*. North-Holland.

Sen, A. K. (1988). Freedom of choice: Concept and content. *European Economic Review*, 32:269–94.

Sen, A. K. (1990). Welfare, freedom and choice: A reply. *Recherches Economiques de Louvain*, 56:451–85.

Sen, A. K. (2002). *Rationality and Freedom*. Harvard University Press., Cambridge, MA.

Sher, I. (2018). Evaluating allocations of freedom. *The Economic Journal*, 128:65–94.

Stantcheva, S. (2023). How to run surveys: A guide to creating your own identifying variation and revealing the invisible. *Annual Review of Economics, 15.*, 15:205–234.

Sugden, R. (1998). The metric of opportunity. *Economics and Philosophy*, 14:307–37.

Sugden, R. (2004). The opportunity criterion: Consumer sovereignty without the assumption of coherent preferences. *The American Economic Review*, 94(4):1014–33.

Toussaert, S. (2018). Eliciting temptation and self-control through menu choices: A lab experiment. *Econometrica*, 86(3):859–89.

UK Government (2023). Patients given more choice to help cut nhs waiting times - what you need to know. https://healthmedia.blog.gov.uk/2023/05/25/patients-given-more-choice-to-help-cut-nhs-waiting-times-what-you-need-to-know/.

Van Hees, M. (1998). On the analysis of negative freedom. *Theory and Decision*, 45:175–97.

World Values Survey (2022). Inglehart–welzel cultural map. https://www.worldvaluessurvey.org/WVSContents.jsp?CMSID=Findings.

Yaari, M. E. and Bar-Hillel, M. (1984). On dividing justly. *Social Choice and Welfare*, 1(1):1–24.

ONLINE APPENDIX

Freedom counts: Cross-country empirical evidence

João V. Ferreira, Nobuyuki Hanaki, Fabrice Le Lec, Erik Schokkaert, Benoît Tarroux

Table of Contents

A. THEORETICAL RANKING RULES

Cardinality family	Cardinality, Weak cardinality, Lex Cardinality-IU, Lex Cardinality-RoO, Diversity
IU family	IU, Lex IU-Cardinality, MaxMax $(0,1)$, MaxMax $(0.5,0.5)$, MaxMax $(1,0)$
Potential preferences family	Range of opportunity, Effective Freedom, Lex RoO-Cardinality
Other rules	Significant choice, MaxAverage $(0,1)$, MaxAverage $(0.5,0.5)$, MaxAverage $(1,0)$, MaxMin $(0,1)$, MaxMin $(0.5,0.5)$, MaxMin $(1,0)$, Choice aversion, Intersection, Trivial

Table A.1. Theoretical ranking rules and families

In this appendix, we briefly describe the theoretical ranking rules we consider in our analysis. Table A.1 lists the rules and their families, while Table A.2 presents their theoretical response patterns.[21] Before describing the rules, we introduce some notation.

Notation

In this appendix, we will use $\succsim$ to refer to either $\succsim^W$ or $\succsim^F$. We use $\#A$ to denote the number of elements in set A.

To define diversity, let S denote a reflexive, symmetric and transitive similarity relation over X, such that xSy is read as "x is similar to y" and $\neg xSy$ is read as "x is not similar to y". A similarity based partition of set A, denoted $\phi(A)$, is defined as a class $\{A_1,...,A_m\}$ such that: (1) $A_1,...,A_m$ are all non-empty subsets of X; (2) $A_1 \cup ... \cup A_m = A$; (3) $A_1,...,A_m$ are pairwise disjoint; and (4) for all $k \in 1,...;,m$, A_k is homogeneous. In our empirical setting, we consider two hospitals similar if they differ at most 1 point in any of the two attributes (e.g. (71,89) and (70,90) in s5 are considered "similar").

To formally describe MaxMax, MaxMin, and MaxAverage rules, we define $q(x_1,x_2) = w_1x_1 + w_2x_2$, where x_j is the quality of x in attribute j and w_j is a quality weight. In our empirical setting, we use the following three weight configurations: $(w_1,w_2) = (0,1)$, $(w_1,w_2) = (1,0)$, and $(w_1,w_2) = (0.5,0.5)$.

[21] Some of these ranking rules have been fully characterized while others have not. Nonetheless, the observable implications of the rules described here are sufficient to determine a theoretical response pattern in our 15 set comparisons. Please refer to the references for formal definitions.

To describe the potential preferences family of rules, let $\mathscr{R} = \{R_1, R_2, ..., R_n\}$ be the set of all *potential preferences*. We consider increasing preferences as potential preferences: if x and y are two alternatives such that $x_i \geq y_i$ for all i, then xR_iy. An alternative x is said to be eligible in a set A if there is at least one potential preference in $\mathscr{R}$ such that x is strictly preferred to all the alternatives in A. If, on the contrary, there is no R_i in $\mathscr{R}$ such that xP_iy for all $y \in A$, then x is said to be ineligible. Denote by $E(A)$ the set of the eligible options in A and by A^B all options $x \in A$ such that no potential preference considers x to be at least as good as all the elements of B.

Cardinality family

– **Cardinality** (Pattanaik and Xu 1990): $A \succsim B$ if and only if $\#A \geq \#B$.

– **Weak cardinality** (see Puppe 1996 for a related rule): $A \sim B$ if $\#A = \#B$ and $A \succsim B$ if $\#A > \#B$.

– **Diversity** (Pattanaik and Xu 2000; see also Gravel 2008a): $A \succsim B$ if and only if $\#\phi(A) \geq \#\phi(B)$.

– **Lex Rule 1 - Rule 2** (Bossert et al. 1994): $A \succ B$ if (i) $A \succ B$ according to Rule 1 or if (ii) $A \sim B$ according to Rule 1 and $A \succ B$ according to Rule 2, where “Rule 1” and “Rule 2” can be cardinality, IU, or RoO.

Indirect utility family

– **Indirect utility** (Arrow 1995; Foster 2011): $A \succsim B$ if and only if $\max(A) R_i \max(B)$, where R_i is i’s actual preference ordering.

– **MaxMax** (Foster 2011): $A \succsim B$ if and only if $\max_{x \in A}\{q(x_1, x_2)\} \geq \max_{y \in B}\{q(y_1, y_2)\}$.

Potential preferences family

– **Range of Opportunity** (Pattanaik and Xu 1998): $A \succsim B$ if and only if $\#[E(A) \setminus A^B] \geq \#[E(B) \setminus B^A]$.

– **Effective freedom** (Foster 2011; see also Arrow 1995 and Sen 2002): $A \succsim B$ if and only if $\max(A) R_i \max(B)$ for all $R_i \in \mathscr{R}$. Otherwise, incomparable.

Other rules

– **Significant choice** (Pattanaik and Xu 2000; see also Sugden 1998): $A \succsim B$ if and only if $\max(\#[\phi(A) \setminus A^B], 1) \geq \max(\#[\phi(B) \setminus B^A], 1)$.

– **MaxMin** (Foster 2011): $A \succsim B$ if and only if $\min_{x \in A}\{q(x_1, x_2)\} \geq \min_{y \in B}\{q(y_1, y_2)\}$.

– **MaxAverage** (Foster 2011): $A \succsim B$ if and only if $\frac{1}{\#A}\sum_{x \in A} q(x_1, x_2) \geq \frac{1}{\#B}\sum_{y \in B} q(y_1, y_2)$.

– **Choice aversion**: (i) $A \cup \{x\} \prec A$ if $\max(A) P_i\, x$ and $A \cup \{x\} \succ A$ otherwise; (ii) for any A and B such that $\#A = \#B$, $A \succsim B$ if and only if $\max(A) R_i \max(B)$.

– **Intersection** (Bossert et al. 1994): $A \succsim B$ if (i) $A \succsim B$ according to cardinality and (ii) $A \succsim B$ according to IU. Otherwise, incomparable.

– **Trivial** (Foster 2011): $A \sim B$ for all A and B.

	Cardinality family					IU family					Potential pref.			Other rules									
	Cardinality	Weak cardinality	Lex Cardinality-RoO	Lex Cardinality-IU	Diversity	Indirect utility	Lex IU-Cardinality	MaxMax (0,1)	MaxMax (0.5,0.5)	MaxMax (1,0)	Range of opportunity	Lex RoO-Cardinality	Effective freedom	Significant choice	MaxMin (0,1)	MaxMin (0.5,0.5)	MaxMin (1,0)	MaxAverage (0,1)	MaxAverage (0.5,0.5)	MaxAverage (1,0)	Choice aversion	Intersection	Trivial
s1	∼	∼	∼	A/B/∼	∼	A/B/∼	A/B/∼	B	∼	A	∼	∼	nc	∼	B	∼	A	B	∼	A	A/B/∼	∼/nc	∼
s2	∼	∼	A	A	∼	A	A	A	A	A	A	A	A	∼	A	A	A	A	A	A	A	nc	∼
s3	∼	∼	A	A	∼	A	A	A	A	A	A	A	A	∼	A	A	A	A	A	A	A	nc	∼
s4	B	B/∼	B	B	B	B/∼	B	B	B	∼	B	B	B/∼	B	∼	∼	A	B	B	A	A/B	B/nc	∼
s5	B	B/∼	B	B	∼	B/∼	B	∼	∼	∼	B	B	B/∼	∼	∼	∼	∼	∼	∼	∼	A/B	B/nc	∼
s6	B	B/∼	B	B	B	B/∼	B	∼	∼	B	B	B	B/∼	B	A	A	∼	A	A	B	A/B	B/nc	∼
s7	B	B/∼	B	B	B	∼	B	∼	∼	∼	∼	B	∼	∼	A	A	A	A	A	A	A	nc	∼
s8	B	B/∼	B	B	B	∼	B	∼	∼	∼	∼	B	∼	∼	A	A	A	A	A	A	A	nc	∼
s9	B	B/∼	B	B	B	∼	B	∼	∼	∼	∼	B	∼	∼	A	A	A	A	A	A	A	nc	∼
s10	B	B/∼	B	B	B	∼	B	∼	∼	∼	∼	B	∼	∼	∼	∼	A	B	A	A	A	nc	∼
s11	A	A/∼	A	A	A	B	B	B	B	B	B	B	B	A	B	B	B	B	B	B	B	nc	∼
s12	A	A/∼	A	A	A	B	B	B	B	B	B	B	B	A	B	B	B	B	B	B	B	nc	∼
s13	A	A/∼	A	A	A	B	B	B	B	B	B	B	B	A	B	B	B	B	B	B	B	nc	∼
s14	∼	∼	∼	A/B/∼	∼	A/B/∼	A/B/∼	B	∼	∼	∼	∼	nc	B	∼	∼	A	B	∼	A	A/B/∼	∼/nc	∼
s15	∼	∼	∼	∼	∼	∼	∼	∼	∼	∼	∼	∼	∼	∼	∼	∼	A	B	∼	A	∼	∼	∼

Notes: Theoretical response patterns respect dominance relations. For rules depending on individual preferences over options that show more than one prediction in the table (e.g. IU in s1), the ranking takes only one of the shown predictions depending on participants' responses to which hospital they prefer (Q2). "nc" denotes that the two menus are not comparable according to that rule.

Table A.2. Theoretical response patterns: All rules

B. Bayes classifications technical details

In this appendix, we present the technical details underlying our Bayes classifications. To calculate posterior probabilities, we use the following expression:

$$P(R_j,\varepsilon_j|c_i) = \frac{P(c_i|R_j,\varepsilon_j)\mu}{P(c_i)}$$

where μ is the prior probability and $P(c_i) = \sum_j \frac{1}{J}\int_0^1 P(c_i|R_j,\varepsilon_j)d\varepsilon_j$.

Following Costa-Gomes et al. (2001) and Ambuehl and Bernheim (2021), we assume that (a) subjects follow their assigned rule with probability $1-\varepsilon_j$ and uniformly randomize over the three possible choices $\{A,B,I\}$ (preference for menu A, preference for B, or indifference) with probability ε_j, where ε_j is distributed uniformly over $[0,1]$; (b) when a rule is irresolute (i.e., it provides an incomplete ranking over two menus, as it is sometimes the case for the effective freedom and intersection rules), the subject randomizes uniformly over the three possible choices; (c) errors are independent across profiles; and (d) the prior distribution over rules is uniform.

For each binary set comparison s, a rule R_j prescribes a subset $Z^s_{R_j} \subseteq \{A,B,I\}$. For each s, subject i makes a choice $c^s_i \in \{A,B,I\}$. We can thus write:

$$\begin{aligned}
P\big(c^s_i = C \mid Z^s_{R_j} = \{C\},\varepsilon_j\big) &= 1-\frac{2}{3}\varepsilon_j \\
P\big(c^s_i = C \mid Z^s_{R_j} = \{D\},\varepsilon_j\big) &= \frac{1}{3}\varepsilon_j \\
P\big(c^s_i = C \mid Z^s_{R_j} = \{C,D\},\varepsilon_j\big) &= \frac{1}{2}-\frac{1}{6}\varepsilon_j \\
P\big(c^s_i = E \mid Z^s_{R_j} = \{C,D\},\varepsilon_j\big) &= \frac{1}{3}\varepsilon_j \\
P\big(c^s_i = C \mid Z^s_{R_j} = \{C,D,E\},\varepsilon_j\big) &= \frac{1}{3}
\end{aligned}$$

where $\{C,D,E\}$ can take values $\{A,B,I\}$. Then, $P(c_i|R_j,\varepsilon_j)$ can be written as follows:

$$P(c_i|R_j,\varepsilon_j) = \prod_{s=1}^{S} P\big(c_i^s|Z_{R_j}^s,\varepsilon_j\big)$$

Finally, we assign a subject i to rule R_j and error probability ε_j^* if and only if:

$$\begin{aligned} & P(R_j,\varepsilon_j^*|c_i) > P(R_k,\varepsilon_k^*|c_i) \quad \forall \quad k \in \{1,...,J\} \setminus j \\ \Leftrightarrow \quad & P(c_i|R_j,\varepsilon_j^*)\frac{\mu_j}{P(c_i)} > P(c_i|R_k,\varepsilon_k^*)\frac{\mu_k}{P(c_i)} \quad \forall \quad k \in \{1,...,J\} \setminus j \end{aligned}$$

As there is no a priori reason to state that $\mu_j \neq \mu_k$, this is equivalent to:

$$P(c_i|R_j,\varepsilon_j^*) > P(c_i|R_k,\varepsilon_k^*) \quad \forall \quad k \in \{1,...,J\} \setminus j$$

C. Sample characteristics and attrition

The observed characteristics of our sample are displayed in Table C.1. Our sample is representative of each country in terms of age group and gender compositions, which means we have roughly half female participants and an average age of 43 in our full sample. As shown in the table, the full sample has also a large share of participants with different perceived social statuses, an overwhelming majority of participants (89%) with upper secondary education or higher (2 and above in the table), a mix of occupations across the population (with a majority of participants either employed or self-employed, 1 and 2 in the table respectively), and perceived health status of 2.71 (where 1 is excellent and 5 is poor). There is heterogeneity between countries among most of these characteristics, which we control for in our cross-country analysis.

The table also shows that the survey was considered fairly easy to understand by our participants. On a scale from 1 (Very difficult) to 10 (Very easy), the mean score is 7.62 for the full sample. The only exception is Japan, where the mean score is 4.76. This evidence, coupled with the relatively high level of education of our sample, supports premises concerning the understanding of the instructions and tasks. Our survey is also short in duration and has only 15 main "tasks", and previous research suggests that response quality in online survey-based studies only degrades after 30 tasks (Bansak et al. 2018).

In terms of attrition, we had a total of 6,347 prospective participants who passed the quota filters.[22] 1,445 (23%) of these were excluded. Among the latter, 98 were excluded because they failed the attention check, 286 passed the attention check but did not finish the survey, and the remaining 1,061 participants stopped the survey before facing the attention check. In fact, the majority of participants who dropped out did not reach the first set comparison (72%). In terms of observed characteristics, we find no difference in age and gender compositions between our final sample and excluded participants.[23] The attrition rate is different across countries, with Japanese and US participants most likely to drop out/be excluded (about 29%), and Chinese and Portuguese the least likely (about 14%). Note that since our main research design does not rely on multiple treatments, differential attrition is not an issue in our setting. However, the differences in attrition rates across countries (and potential differences in terms of attrition characteristics across countries), are among the reasons why we control for observed characteristics in our cross-country analysis.

[22]This figure does not include participants who were excluded because of speeding, for whom the survey company did not provide us data.

[23]Note that we asked about age and gender at the beginning of the survey for quota purposes, while we asked about the remaining socio-demographics at the end of the survey. On the one hand, this has the drawback that we do not have data on the latter socio-demographics for excluded participants. On the other hand, this design choice decreases noise since it lowers participants' cognitive load and survey fatigue during our main tasks.

	Age (mean)	Female (fraction)	Social status 1 (fraction)	Social status 2	Social status 3	Education 1 (fraction)	Education 2	Education 3	Education 4	Education 5	Occupation 1 (fraction)	Occupation 2	Occupation 3	Occupation 4	Occupation 5	Occupation 6	Health (mean)	Difficulty (mean)
All	42.59	0.51	0.31	0.42	0.27	0.11	0.24	0.13	0.40	0.12	0.53	0.10	0.11	0.13	0.09	0.04	2.71	7.62
BR	37.20	0.51	0.26	0.39	0.35	0.35	0.06	0.08	0.41	0.09	0.46	0.24	0.13	0.06	0.10	0.02	2.48	8.12
CH	39.43	0.49	0.31	0.49	0.20	0.01	0.07	0.07	0.76	0.09	0.71	0.04	0.00	0.13	0.10	0.01	2.53	8.30
CO	34.39	0.52	0.32	0.39	0.29	0.05	0.29	0.21	0.35	0.10	0.45	0.22	0.14	0.02	0.15	0.02	2.36	7.87
FR	47.20	0.52	0.31	0.48	0.21	0.07	0.33	0.17	0.26	0.18	0.53	0.04	0.05	0.24	0.06	0.07	2.83	7.27
JA	47.62	0.51	0.22	0.40	0.38	0.04	0.30	0.19	0.40	0.06	0.57	0.07	0.23	0.05	0.04	0.04	3.38	4.76
NL	46.50	0.50	0.44	0.37	0.19	0.23	0.33	0.05	0.25	0.14	0.48	0.09	0.10	0.17	0.08	0.10	2.78	7.62
PT	40.00	0.52	0.23	0.51	0.26	0.02	0.29	0.11	0.38	0.19	0.64	0.11	0.07	0.06	0.10	0.02	2.79	8.28
TR	38.13	0.50	0.54	0.36	0.10	0.03	0.21	0.03	0.59	0.15	0.55	0.06	0.09	0.12	0.14	0.05	2.54	7.61
UK	48.58	0.50	0.17	0.42	0.41	0.09	0.28	0.20	0.29	0.13	0.49	0.06	0.10	0.25	0.03	0.07	2.76	8.21
US	46.85	0.52	0.28	0.38	0.34	0.18	0.22	0.20	0.29	0.11	0.39	0.05	0.18	0.25	0.09	0.04	2.65	8.23

Notes: *Social status* shows perceived status of 1 (Upper and Upper middle class), 2 (Lower middle class), or 3 (Working class and Lower class). *Education* shows completed level of 1 (Lower secondary education), 2 (Upper secondary education), 3 (Post-secondary non-tertiary education), 4 (Bachelor's or equivalent level), or 5 (Master's or equivalent level and higher). *Occupation* shows current occupation of 1 (Employed), 2 (Self-employed), 3 (Unemployed), 4 (Retired), 5 (Student), and 6 (Inactive). *Health* shows perceived health status from 1 (Excellent) to 5 (Poor). *Difficulty* shows perceived difficulty of the survey from 1 (Very difficult) to 10 (Very easy).

Table C.1. Sample characteristics

D. Robustness checks: Main classifications

In this appendix, we perform several robustness tests of our classification results. We remove inattentive participants (Section D.1), test the results per total response time (Section D.2), and test for potential effects of the versions of the survey (Section D.3). We also compare our classification results to a random benchmark of 5,000 artificial subjects (Section D.4), and perform a cluster analysis to check if we omitted empirically important rules in our classifications (Section D.5). Finally, we check if our results hold for the set comparisons where it is easy to infer Mr. Green and Mr. Yellow's preferences over alternatives (Section D.6).

D.1 Inattentive participants

We present here the robustness results for the Naive Bayesian classifier by removing inattentive subjects. To do that, we use the answers to the IU question (Q2) for set comparisons where there is a clear dominant (or dominated) option. We focus on set comparisons where the subjects can easily identify the dominant or dominated alternatives (s3, s8, s9, and s13).[24] This approach provides us with a clear inbuilt test of attention beyond our basic screening via the attention check and total response time. Accordingly, an attentive subject should respond as follows:

1. (80,80) strictly preferred to (60,60) in set comparison s3.
2. (70,90) or (71,89) strictly preferred to (60,60) in s8.
3. (70,90) or (71,89) strictly preferred to (20,30) in s9.
4. (81,92) strictly preferred to (70,70), (50,71) and (72,65) in s13.

With *filter 1*, we only keep participants who respect all of the above conditions, while with *filter 2* we keep participants who respect at least three out of four of the above

[24]We do this because (i) in some cases the dominant or dominated alternative is not salient and (ii) participants may be indifferent between a dominant/dominated alternative and a "similar" non-dominant/dominated alternative if they have a "thick" indifference curve.

conditions (i.e., we allow participants to make at most one mistake).[25] Table D.1 shows the number of subjects per attention filter.

	Filter 1 (no error)	Filter 2 (at most 1 error)	All subjects
#	3,952	4,650	4902
%	80.62	94.86	100

Table D.1. Number of subjects per attention filter

Figure D.1 reports the classification results (distribution of ranking rules) for the baseline (all subjects) and the two attention filters. The results with the two filters are very similar to our main results for both the freedom and welfare classifications. In addition, these results show that the two main groups for each question (cardinality and IU respectively) are not driven by inattentive participants. In fact, the percentages increase with stricter filters of attention. This suggests that cardinality and IU are not being used as simple heuristics to quickly answer the questions. It is also worth noting that (not surprisingly) the trivial type increases as inattention increases, suggesting that this is capturing a small percentage of subjects who are using this ranking rule as a simple heuristic to respond.

To test the robustness of these results, we regress the probability of endorsing the main rules and families for different numbers of errors and controlling for country and individual observed characteristics (gender, age, social status, education, occupation, and health status). Table D.2 reports the estimates of the average marginal effects of the number of errors, using logistic regressions for main rules and families. When looking at the freedom classification, we find that subjects are statistically significantly less likely to be assigned to the cardinality rule (and the cardinality family) as the number of errors increases. For the welfare classification, we also find that subjects are statistically significantly less likely to be assigned to the IU rule (and the IU family) as the number of errors increases.

[25] Almost all participants (98.49%) respect at least two out of the four above conditions. Therefore, it does not make sense to have a filter that allows participants to make at most two mistakes (or more).

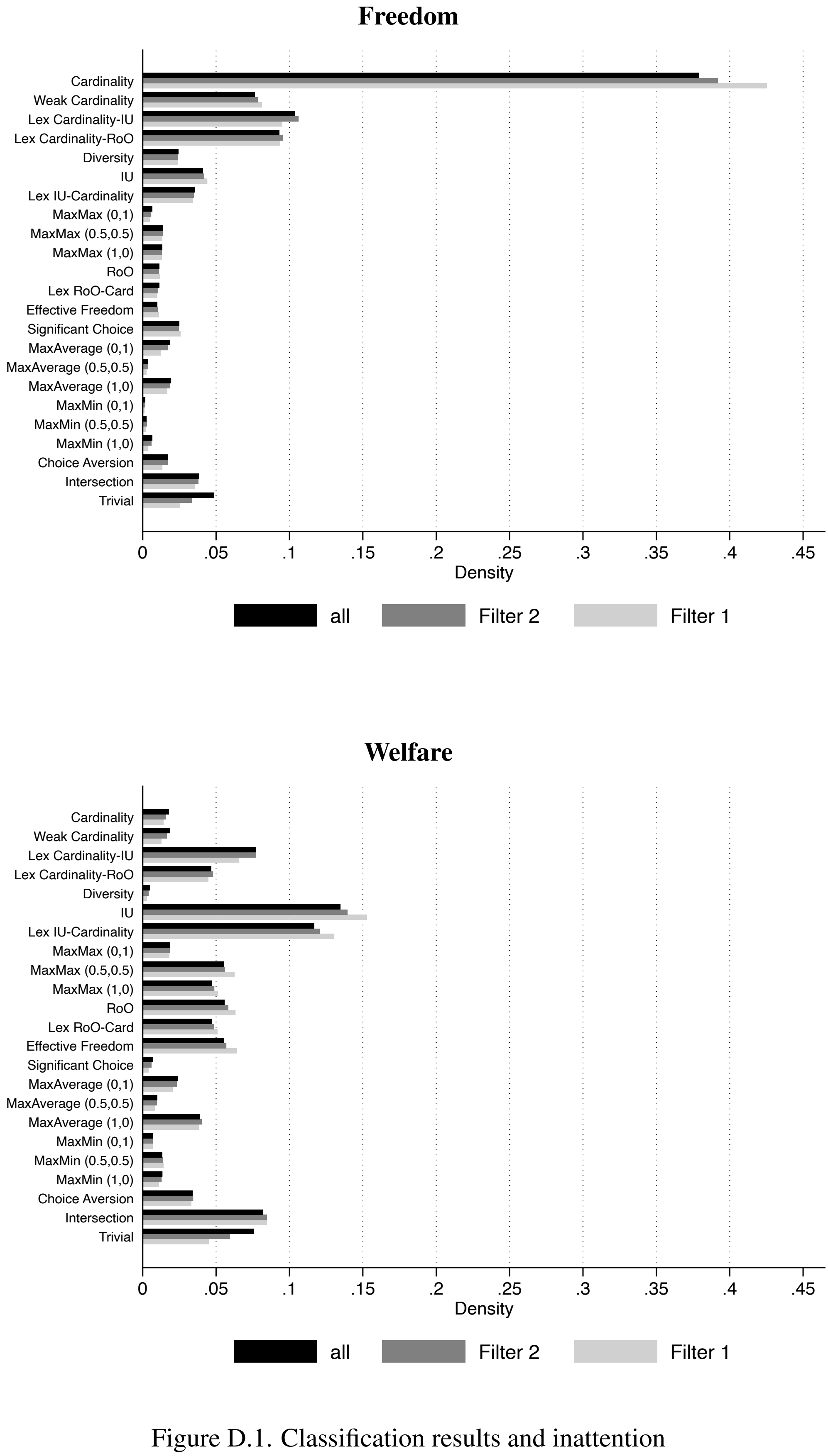


Figure D.1. Classification results and inattention

	Freedom classification		Welfare classification	
	Cardinality	Cardinality family	IU	IU family
# errors = 0	(ref)	(ref)	(ref)	(ref)
# errors = 1	-0.205***	-0.136***	-0.0906***	-0.216***
	(0.0175)	(0.0195)	(0.0107)	(0.0170)
# errors ≥ 2	-0.268***	-0.379***	-0.108***	-0.240***
	(0.0245)	(0.0305)	(0.0146)	(0.0256)
Wald tests (*p*-value)				
# errors = 1 vs # errors ≥ 2	0.0233	<0.001	0.2841	0.3975

Notes: This table reports the average marginal effects of the number of errors using a logit regression for the main rules/families in the freedom and welfare classifications, controlling for country and observed characteristics.

Table D.2. Main classification results and number of errors: Regression results

D.2 Total response time

We can also test how our results relate to participants' response times. This gives us an additional test of attention. The mean and median response times were 799 seconds (≈ 13 minutes) and 661 seconds (≈ 11 minutes) respectively. For this analysis, we divide the sample into four quartiles of total response time. As shown in Table D.3, Quartile 1 corresponds to the fastest participants (median response time ≈ 7 minutes) and Quartile 4 corresponds to the slowest participants (median response time ≈ 21 minutes).

	Quartile 1	Quartile 2	Quartile 3	Quartile 4	All
Mean	400	575	787	1,889	799
Median	403	577	780	1,276	661

Table D.3. Total response time in seconds

Figure D.2 reports the classification results per quartile. When focusing on freedom, it is clear that the fastest participants (Quartile 1) are less likely to respond in line with the cardinality rule and other cardinality family ranking rules. This accords with our previous result about (in)attention. It is also noticeable that the fastest participants are more likely to endorse IU and trivial rules for the freedom classification. When looking at the welfare classification, the fastest participants are less likely to endorse IU, other IU family rules (e.g. Lex IU-cardinality), and more likely to endorse the trivial rule. These

	Freedom classification		Welfare classification	
	Cardinality	Cardinality family	IU	IU family
Quartile 1 ($Q1$)	(ref)	(ref)	(ref)	(ref)
Quartile 2 ($Q2$)	0.137***	0.176***	0.0158	0.0749***
	(0.0179)	(0.0191)	(0.0141)	(0.0195)
Quartile 3 ($Q3$)	0.212***	0.231***	0.00433	0.0554***
	(0.0185)	(0.0188)	(0.0140)	(0.0197)
Quartile 4 ($Q4$)	0.192***	0.203***	-0.00478	0.0475**
	(0.0189)	(0.0194)	(0.0142)	(0.0202)
Wald tests (p-value)				
$Q2 = Q3$	<0.001	0.0015	0.4142	0.3255
$Q2 = Q4$	0.0053	0.1269	0.1443	0.1726
$Q3 = Q4$	0.3207	0.1013	0.4991	0.6860

Notes: This table reports the average marginal effects of the quartiles of total response time using a logit regression for the main rules/families in the freedom and welfare classifications, controlling for country and observed characteristics.

Table D.4. Main classification results and total response time: Regression results

results are again aligned with the ones from the previous section, suggesting that the cardinality family rules for freedom and IU family rules for welfare are not being used as simple heuristics.

To test the robustness of these raw results, we regress the probability of endorsing the main rules and families for different quartiles using a logistic model (Table D.4). The regression results confirm the previous findings. The fastest participants (Quartile 1) are statistically significantly less likely to endorse the cardinality rule and the cardinality family for the Freedom question. For the Welfare question, the fastest participants are statistically significantly less likely to endorse the IU family than other participants.[26]

D.3 Version of the survey

Subjects face one of two versions of the survey (between-subject treatments). In version *FreedomIU*, the Freedom question appears first and the IU question appears second, while in version *IUFreedom* this order is reversed. Figure D.3 reports the classification results per version of the survey. The most noticeable difference is the prevalence of the

[26]These effects are not always monotonic, but non-monotonic differences are only statistically significant at 10% between Quartiles 3 and 4 for the cardinality family and Quartiles 2 and 4 for the IU family.

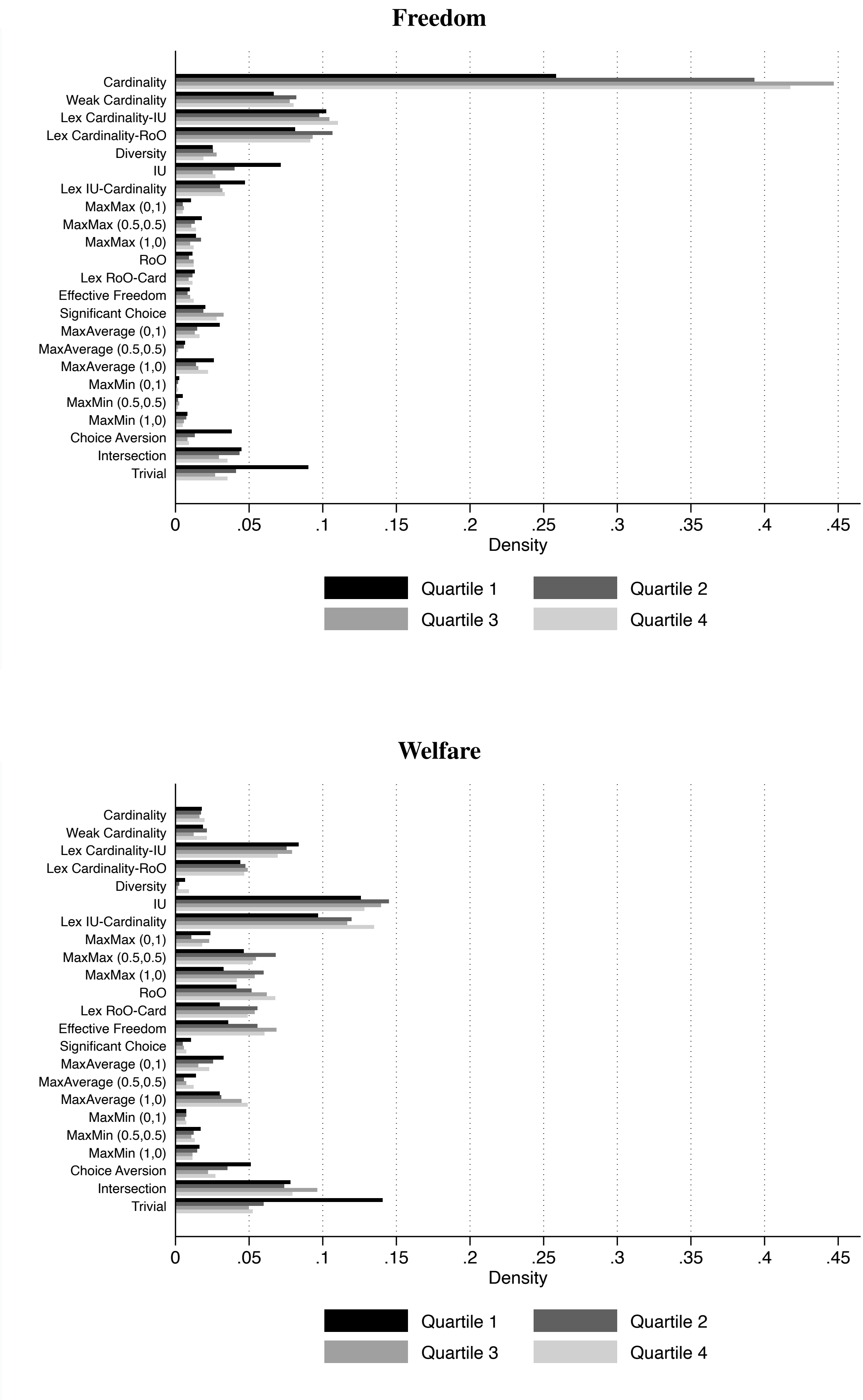


Figure D.2. Classification results and total response time

	Freedom classification		Welfare classification	
	Cardinality	Cardinality family	IU	IU family
FreedomIU	(ref)	(ref)	(ref)	(ref)
IUFreedom	-0.0723***	-0.0825***	0.0112	-0.00103
	(0.0135)	(0.0129)	(0.00970)	(0.0138)

Notes: This table reports the average marginal effects of the version of the survey using a logit regression for the main rules/families in the freedom and welfare classifications, controlling for country and observed characteristics.

Table D.5. Main classification results and version of the survey: Regression results

	Average	**Cardinality family**	Cardinality	Weak cardinality	Lex Cardinality-IU	Lex Cardinality-RoO	Diversity	**IU family**	Indirect Utility	Lex IU-Cardinality	MaxMax (0,1)	MaxMax (0.5,0.5)	MaxMax (1,0)	**Potential Pref. family**	**Other rules**
Random answers	0.70	0.69	0.73	0.61	0.73	0.69	0.71	0.70	0.69	0.71	0.72	0.65	0.71	0.66	0.71

Notes: This table reports the average error (ε^*) for main ranking rules and families with random answers. "Cardinality family", "IU family", and "Potential Pref. family" report the average error for the rules in the respective families, while "Other rules" reports the average error for the theoretical rules that do not belong to these families.

Table D.6. Goodness-of-Fit of the Bayesian classifier for random answers

cardinality rule in the two versions (42% in FreedomIU and 34% in IUFreedom). As shown in Table D.5, this difference is statistically significant. This suggests an effect of the order of the main questions. However, this effect is small and — as shown in Figure D.3 — is not driving our results. For the main rules and family in the welfare classification, the effect is smaller and it is not statistically significant.

D.4 Random benchmark

To evaluate our classification results, we generate 5,000 artificial subjects whose simulated answers are uniformly distributed across options in their ranking over menus and in their ranking of hospitals. Results are reported in Figure D.4. The distribution of rules clearly differs from the one of the real subjects. For this random benchmark, no rule gathers more than 9% of artificial subjects and the average noise ε^* is 0.7, which is much higher than the one for real participants for the Freedom (0.21) and Welfare (0.30) questions.

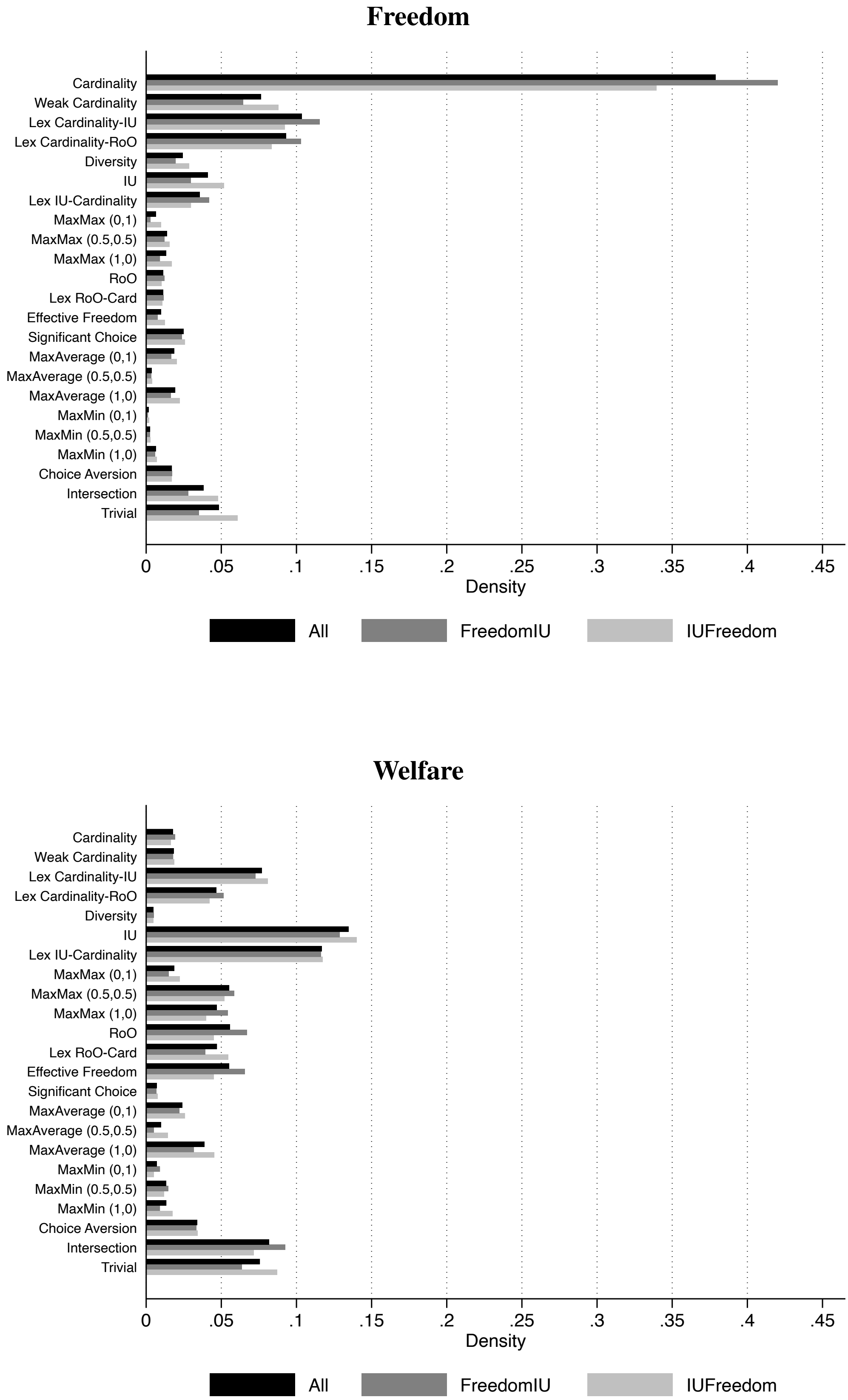


Figure D.3. Classification results and version of the survey

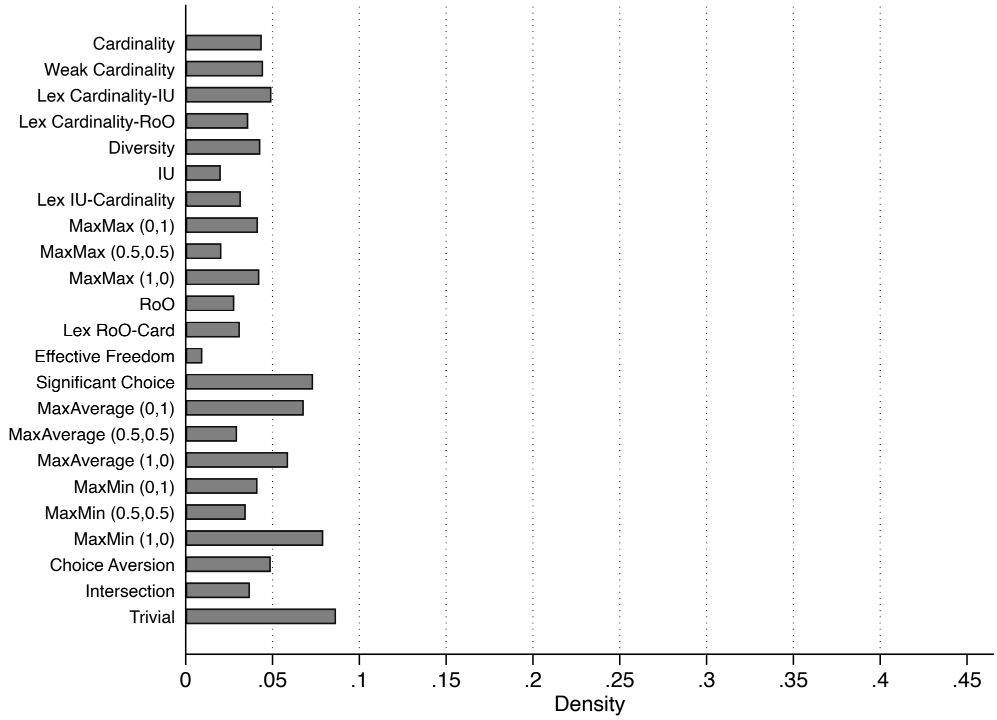


Figure D.4. Bayesian classification for Random answers

D.5 Cluster analysis

The Bayesian classifier is a powerful method to determine subjects' "types" and estimate the frequency with which our pre-specified theoretical rules are implicitly followed. However, this method is only valid as long as the pre-specified rules are exhaustive. An omitted rule could bias the results. Following Costa-Gomes and Crawford (2006) and Ambuehl and Bernheim (2021), we perform various k-modes cluster analyses to assess whether this is the case. In other words, we look for patterns in the responses of subjects that could not be accounted for by any of the theoretical rules we used in our Bayesian classifier. A specific challenge arises from the fact that the rankings of opportunity sets depend, for "preference-based rules", on the underlying preferences over alternatives. Therefore, six of our preference-based rules do not have a single possible response vector for the 15 binary set comparisons.[27] To take that into account, we proceed as follows:

[27]The rules are IU, Lex Cardinality-IU, Lex IU-Cardinality, effective freedom, choice aversion, and intersection. See Table A.2 for the set comparisons for which these rules do not have a single prediction.

(1) For each participant and each theoretical preference-based rule, we compute the response vector for the 15 set comparisons that the rule implies using the participant's preferences over alternatives/hospitals. For each theoretical rule, we can then define the set of all response vectors that are possible given the preferences over alternatives of our participants. For instance, IU is compatible with different response vectors depending on individual preferences over alternatives on s1, s4 to s6, and s14, and we calculate which ones are possible with the preferences over alternatives of our participants. (2) We run cluster analyses and obtain different clusters of subjects. We describe these clusters by the vector of modal answers for the 15 set comparisons (hereafter the *modal vector of a cluster*). If the modal vector of a cluster is the same as one of the possible response vectors of one of our pre-specified theoretical rules, we then identify the cluster to this rule. (3) Given that the cluster procedure is exploratory, we report results for different numbers of estimated clusters. We show the frequency of each cluster and the modal vector of the clusters that do not correspond to our theoretical rules. If two different clusters have the same modal vector, we add their frequencies (i.e., join them in one cluster for presentation purposes).

Freedom question. First, as shown in Table D.7, the clustering method confirms the main result of the naive Bayesian classifier for the Freedom classification, with a large proportion of subjects being classified as cardinality and Lex Cardinality-IU. Second, the results clearly show that no particularly frequent and robust rule has been omitted. As we increase the number of clusters, behavior seems to diverge into many different patterns, not hinting at any particularly salient missing rule. As seen in the table, there are only two clusters that do not correspond to our pre-specified rules that are moderately *frequent* with 12 or 20 clusters (Other 12 and Other 16).[28] Other 12 can be described as a variant of Lex Cardinality-IU, where subjects exhibit choice aversion, particularly when adding dominated options as in set comparisons s7, s8, and s9. Other 16 has a modal vector very similar to the cardinality rule, exhibiting choice aversion (or indifference to more choice)

[28]By "frequent" we mean more than 5% of subjects, since 5% is the frequency expected for a 20-cluster analysis with random data.

Number of clusters:	6	9	12	20
Cardinality	49.3	44.8	41.8	40.4
IU	9.9	10.2	6.7	5.1
Lex Cardinality-IU		17.0	11.2	7.3
Lex Cardinality-RoO		c(Lex Card-IU)	c(Lex Card-IU)	
MaxMax $(1,0)$	c(IU)			
Trivial	6.4	5.9	5.9	5.7
Other 1 (∼∼∼B∼∼∼∼∼∼∼AAA∼∼)	6.3	6.6		4.5
Other 2 (∼∼∼BBAAA∼BAAA∼∼)	13.3			
Other 3 (AAABBAABBBBBBAB)	14.7			
Other 4 (AAAAA∼AAA∼AAAA∼)		3.3		
Other 5 (AAABBABBBBBBBAB)		6.8		
Other 6 (BAABBAAAAABBBBA)		5.3		
Other 7 (AAABBAABBBAAAAB)			6.3	
Other 8 (∼A∼B∼∼∼AABAAA∼∼)			1.7	
Other 9 (∼∼∼B∼A∼∼∼∼AAA∼∼)			6.7	
Other 10 (AAABBA∼∼∼∼BBBAB)			5.1	
Other 11 (∼∼AB∼A∼AABAAAB∼)			1.7	
Other 12 (AAABBAAAABAAABB)			6.3	5.6
Other 13 (BAABAAAAAABBBBA)			4.9	
Other 14 (AAABBA∼AA∼ABA∼∼)			1.7	
Other 15 (∼AABBAAAABAABB∼)				2.7
Other 16 (∼∼∼BBAA∼∼BAAA∼∼)				6.2
Other 17 (AAABBA∼AB∼ABBB∼)				1.4
Other 18 (AAABBBBBBBAAAB)				2.6
Other 19 (AAAB∼A∼∼∼∼BABA∼)				3.3
Other 20 (AAA∼A∼∼BB∼BBBA∼)				1.2
Other 21 (BAABBAABBBBABBA)				2.5
Other 22 (AAABBAAABBBBBAB)				4.0
Other 23 (AAABBBBBBBABB∼∼)				1.1
Other 24 (BBBAAAAAAABBBBA)				1.6
Other 25 (∼AAB∼∼BBBABBB∼∼)				1.0
Other 26 (∼∼∼BBA∼AABAAA∼∼)				3.8

Notes: Numbers are the proportion of each cluster in percentage. If the modal vector of a cluster corresponds to one of our pre-specified theoretical rules, we name this cluster after this rule. Otherwise, we name the cluster as "Other". "c()" denotes that the rule in the row is compatible with (and non-distinguishable from) the modal vector of the cluster associated with the rule named in parenthesis. Sequence of ∼, A, and B after "Other #" corresponds to the modal pattern for set comparisons s1 to s15 clusters, from left to right. If two different clusters have the same modal vector, we add their frequencies and join them in one cluster.

Table D.7. Cluster Analyses: Freedom question

Number of clusters:	6	9	12	20
IU	29.3	19.3	19.3+(Effective)	13.2
Effective Freedom			10.1	
Lex Cardinality-IU	15.0	11.7		4.4+(Lex Card-RoO)
Lex Cardinality-RoO	c(Lex Card-IU)			6.6
Trivial		9.2		
Other 1 (AAABBAAAA∼BBBAB)	17.6			8.6
Other 2 (AAABBABBB∼BBBA∼)	16.1			
Other 3 (∼AAB∼A∼∼∼∼AAB∼∼)	10.7			
Other 4 (∼AA∼∼∼∼∼∼∼BBB∼∼)	11.3			
Other 5 (∼AAB∼A∼∼∼∼∼AB∼∼)		10.0		
Other 6 (AAAB∼AA∼∼∼BBBA∼)		12.5		
Other 7 (AAABBABBABBABAA)		4.2		
Other 8 (∼AABBAAABBBBBB∼)		9.8		
Other 9 (AAABBABBB∼BBBAB)		13.1		
Other 10 (AAABBAAAA∼BAB∼∼)		10.2		
Other 11 (AAABBA∼∼A∼AABA∼)			4.2	
Other 12 (∼AAABABBB∼BBBBB			3.2	
Other 13 (∼AA∼∼∼∼∼∼∼∼∼B∼∼)			18.6	15.0
Other 14 (AAABBAAAA∼BBBBA)			9.4	5.4
Other 15 (AAABBAABBBBBBAB)			7.8	
Other 16 (∼AAB∼AABBBBBBBA)			3.0	
Other 17 (AAAB∼ABBBABBBAB)			4.7	
Other 18 ∼AABBAAAA∼AAA∼∼)			12.2	
Other 19 (AAABBAAAABBAB∼∼)			7.4	
Other 20 (AAA∼∼∼A∼∼B∼∼∼A∼)				2.2
Other 21 (∼AABB∼AABB∼A∼A∼)				1.1
Other 22 (AAA∼AAA∼∼ABBBA∼)				2.3
Other 23 (AAA∼B∼∼∼B∼BBBB∼)				3.9
Other 24 (∼AABBAA∼∼∼BBBA∼)				6.7
Other 25 (AAAB∼BABB∼BBB∼∼)				2.7
Other 26 (∼AABBA∼AB∼BBB∼∼)				5.5
Other 27 (AAABBAABBBABBAB)				5.9
Other 28 (AAABBAAAB∼AAA∼∼)				4.2
Other 29 (BAABBA∼∼B∼BBB∼A)				3.0
Other 30 (BAAABBBBBABBBBB)				2.9
Other 31 (BAA∼∼AA∼A∼ABBA∼)				1.7
Other 32 (∼AAB∼AA∼∼∼AAA∼∼)				3.8
Other 33 (AAAB∼∼B∼B∼AB∼∼A)				0.9

Notes: Numbers are the proportion of each cluster in percentage. If the modal vector of a cluster corresponds to one of our pre-specified theoretical rules, we name this cluster after this rule. Otherwise, we name the cluster as "Other". "c()" denotes that the rule in the row is compatible with (and non-distinguishable from) the modal vector of the cluster associated with the rule named in parenthesis. "+()" denotes that the cluster associated with the rule in parenthesis is also compatible with the rule in the row; therefore, the percentage of the cluster associated with the rule in parenthesis can be added to the percentage of the cluster in the row to give the total percentage compatible with the rule in the row. Sequence of ∼, A, and B after "Other #" corresponds to the modal pattern for set comparisons s1 to s15 clusters, from left to right. If two different clusters have the same modal vector, we add their frequencies and join them in one cluster.

Table D.8. Cluster Analyses: Welfare question in set comparisons s6 to s9. Both these patterns align with the results reported in our main analysis about the presence of some choice aversion (e.g. that subjects classified as cardinality are more likely to depart from this rule when we add an option of "poor quality"). Overall, this analysis strongly suggests that we have not omitted any important rule in our freedom classification.

Welfare question. Consistent with the Bayesian classification results, we find that response clusters for the Welfare question are more heterogeneous than for the Freedom question. From our theoretical rules, we observe that IU is the most frequent (as in our main results) and that effective freedom, Lex Cardinality-IU, and Lex Cardinality-RoO — rules with some frequency in our Welfare classification — also correspond to clusters when we use 12 or 20 clusters. Not surprisingly, we also find that — as compared with the cluster analysis for the Freedom question — there is a larger proportion of subjects assigned to clusters that do not correspond to our theoretical rules. From these, there are several that have more than 5% of participants, but only two are "robust" by being present with both 12 and 20 clusters (Other 13 and Other 14). Other 13 has a modal vector that corresponds to a relatively simple and plausible quality-based rule: one set is overall better than another when it contains an option that clearly dominates the best option of the other set (as in s2, s3, and s13), otherwise both sets provide the same overall welfare. It is compatible with IU with a thick indifference curve when comparing opportunity sets. Other 14 can be interpreted as IU with choice aversion for set comparisons s6 to s9. Such a rule can accommodate the modal vector of this cluster except for s15. Overall, this cluster analysis is consistent with our main welfare classification results: IU family rules and other quality-based rules are predominantly used to evaluate the welfare of sets, but there is considerable heterogeneity in the particular rules that subjects implicitly follow to compare the welfare of opportunity sets.

In sum, we find little evidence that a major significant rule is missing from our main analysis. Indeed, the patterns most frequently observed that are not captured by any of our theoretical rules, are either very similar to one of our rules with evidence of choice

Question	**Average**	**Cardinality family**	Cardinality	Weak cardinality	Lex Cardinality-IU	**IU family**	Indirect Utility	Lex IU-Cardinality	**Other rules**
Freedom	0.20	0.15	0.11	0.07	0.27	0.30	0.25	0.40	0.28
Welfare	0.26	0.35	0.35	0.24	0.36	0.24	0.21	0.32	0.26

Notes: This table reports the average error (ε^*) for distinguishable rules and families in the restricted set of menu comparisons s2, s3, s7 to s13, and s15.

Table D.9. Goodness-of-Fit of the Bayesian classifier for restricted set comparisons

aversion for particular sets (Other 12 and Other 16 for the Freedom question and Other 14 for the Welfare question) or correspond to a plausible rule that is very similar to one of our main theoretical rules (Other 13 for the Welfare question).

D.6 Restricted set comparisons

In this appendix, we report the classification results for the set comparisons where it is easy to infer Mr. Green and Mr. Yellow's preferences over alternatives (s2, s3, s7 to s13 and s15). To perform the classifications, we merge the rules that have the same predictions (i.e., cannot be distinguished) on this restricted set of menu comparisons: (i) cardinality with diversity, (ii) Lex cardinality-IU with Lex cardinality-RoO, (iii) IU with RoO, MaxMax rules, and effective freedom, (iv) Lex IU-cardinality with RoO-cardinality, (v) MaxAverage (0.5,0.5) and choice aversion, (vi) MaxMin (0,1) with MaxMin (0.5,0.5) and (vii) MaxMin (1,0) with MaxAverage (1,0).

Figures D.5 and D.6 summarize the results for the Freedom and Welfare questions. Results are similar to our main results, with the size-based rules being the most prominent for the freedom evaluation (63% of subjects), and the preference-based rules being the most popular for the welfare evaluation (39% of subjects). Table D.9 also shows that the goodness-of-fit is similar to the one we obtain with the full set of menu comparisons. As in our main results, we find that the goodness-of-fit is better for the freedom classification (0.21) than the welfare classification (0.28).

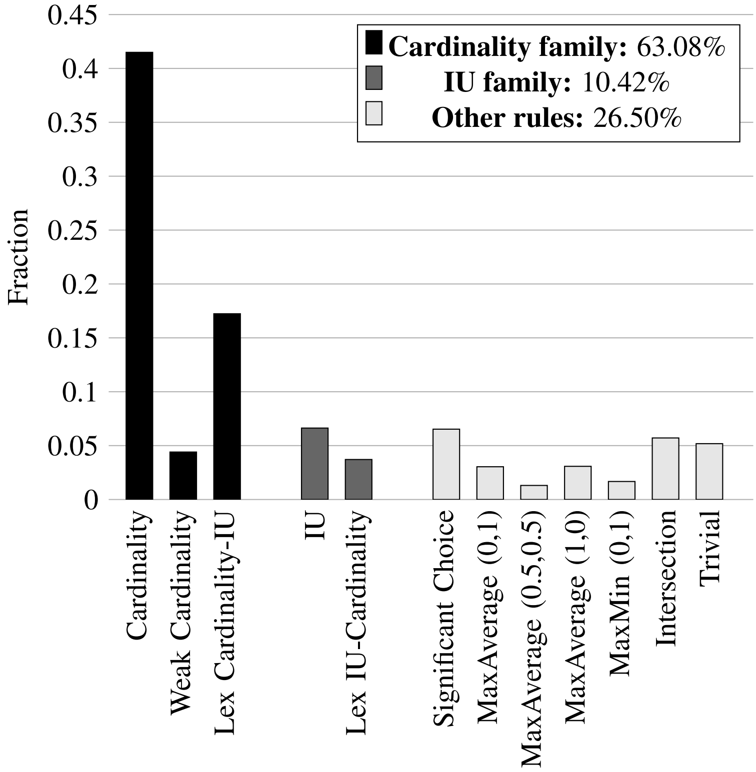


Figure D.5. Bayesian classification for restricted set comparisons (Freedom question)

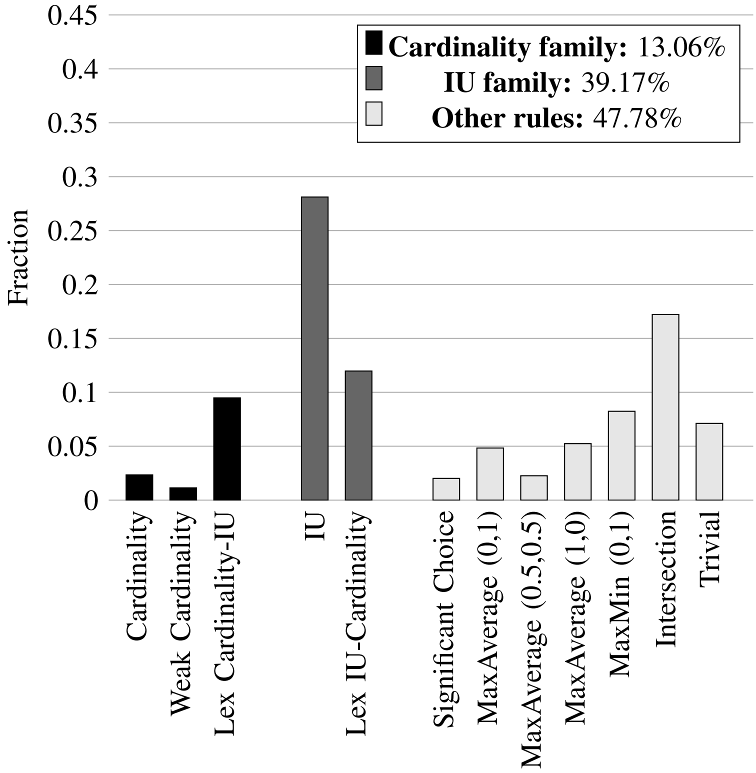


Figure D.6. Bayesian classification for restricted set comparisons (Welfare question)

E. ROBUSTNESS CHECKS: INTRINSIC VALUE OF FREEDOM

In this appendix, we test the robustness of our results for the intrinsic value of freedom when we remove inattentive participants (Section E.1), when we divide the sample for different total response times (Section E.2), and when we look at participants for each of the two versions of the survey (Section E.3).

E.1 Inattentive participants

Table E.1 summarizes our results when we impose the filters described in Appendix D.1. As shown in the table, results are very similar across filters. Still, when we control for the number of "relevant cases" (cases with a conflict between freedom and IU), we see that participants who did not make an error (Filter 1) are statistically significantly less likely to show IvFoC than other participants (see Table E.2). However, the magnitude of this effect is very small. Participants who did not make an error show IvFoC in 46% of relevant cases, which is very similar to the 48% for the full sample.

E.2 Total response time

Results for IvFoC per quartile of total response time are shown in Table E.3. Overall results are again very similar for different quartiles. When controlling for the number of relevant cases, the fastest participants (Quartile 1) are statistically significantly more likely to exhibit IvFoC than other participants (see Table E.4). However, this seems driven by the number of relevant cases and subjects in all quartiles reveal giving intrinsic value to freedom in more than 45% of relevant cases. Taken together with our previous robustness checks (Appendix E.1), this suggests that our IvFoC results are not driven by inattentive participants.

E.3 Version of the survey

Table E.5 presents the results for the intrinsic value of freedom for the two different versions of the survey. We can see that participants in *IUFreedom* version are more

	Relevant cases		IvFoC ($C \succ_i^W D$ in relevant cases)					
	Mean (out of 15)	% subjects > 0	Mean	% of subjects s.t. IvFoC revelations $= 0$	≥ 1	≥ 2	≥ 3	≥ 4
$C \succ_i^F D$ and $\max(D) R_i \max(C)$								
All	7.15	96.43	3.42	15.12	84.88	70.60	57.67	45.84
Filter 1	7.33	97.27	3.36	14.90	85.10	70.42	56.96	44.76
Filter 2	7.25	96.99	3.45	14.43	85.57	71.29	58.32	46.39
$C \succ_i^F D$ and $\max(D) I_i \max(C)$								
All	5.13	95.19	2.87	17.50	82.50	66.67	51.96	38.43
Filter 1	5.17	96.05	2.85	17.05	82.95	66.88	52.18	38.28
Filter 2	5.21	95.85	2.92	16.65	83.35	67.63	53.01	39.35
$C \succ_i^F D$ and $\max(D) P_i \max(C)$								
All	2.02	80.74	0.54	66.10	33.90	14.46	4.92	0.78
Filter 1	2.16	84.92	0.51	67.66	32.34	13.82	4.48	0.33
Filter 2	2.05	81.87	0.53	66.39	33.61	14.17	4.69	0.67

Notes: The first column shows the mean number of responses per subject (in a total of 15 set comparisons) such that C (either set A or B) is considered to have more freedom than D (A if $C = B$ or B if $C = A$) and $\max(D) R_i \max(C)$ (first three rows), $\max(D) I_i \max(C)$ (second three rows), or $\max(D) P_i \max(C)$ (third three rows). The second column shows the percentage of participants for which the number of relevant cases is at least one. The third column shows the mean number of responses per subject such that $C \succ_i^W D$ in the relevant cases (i.e., when $C \succ_i^F D$ and $\max(D) R_i \max(C)$/$\max(D) I_i \max(C)$/$\max(D) P_i \max(C)$). The last five columns present the percentage of subjects who reveal IvFoC among the participants for which the number of relevant cases is at least one.

Table E.1. IvFoC and inattention

	IvFoC ($C \succ_i^W D$ in relevant cases)		
	$C \succ_i^F D$ and $\max(D) R_i \max(C)$	$C \succ_i^F D$ and $\max(D) I_i \max(C)$	$C \succ_i^F D$ and $\max(D) P_i \max(C)$
# errors = 0	(ref)	(ref)	(ref)
# errors = 1	0.838***	0.316***	0.368***
	(0.0917)	(0.0717)	(0.0344)
# errors ≥ 2	0.259*	-0.0121	0.391***
	(0.147)	(0.115)	(0.0538)
Wald tests (p-value)			
# errors = 1 vs # errors ≥ 2	<0.001	0.0111	0.7041

Notes: This table reports the marginal effects of the number of errors using an OLS regression for the three measures of IvFoC, controlling for the number of "relevant cases", country, and observed individual characteristics.

Table E.2. IvFoC and number of errors: Regression results

	Relevant cases		IvFoC ($C \succ_i^W D$ in relevant cases)					
	Mean (out of 15)	% subjects > 0	Mean	% of subjects s.t. IvFoC revelations $= 0$	≥ 1	≥ 2	≥ 3	≥ 4
$C \succ_i^F D$ and $\max(D) R_i \max(C)$								
All	7.15	96.43	3.42	15.12	84.88	70.60	57.67	45.84
Quartile 1	6.32	93.18	3.39	17.30	82.70	67.75	56.05	44.84
Quartile 2	7.21	97.30	3.36	16.46	83.54	69.62	56.10	44.88
Quartile 3	7.62	97.80	3.44	14.11	85.89	71.94	58.48	46.41
Quartile 4	7.48	97.47	3.49	12.58	87.42	73.12	60.05	47.22
$C \succ_i^F D$ and $\max(D) I_i \max(C)$								
All	5.13	95.19	2.87	17.50	82.50	66.67	51.96	38.43
Quartile 1	4.63	91.47	2.77	20.71	79.29	63.04	49.39	36.72
Quartile 2	5.14	95.82	2.86	18.35	81.65	65.77	51.11	38.66
Quartile 3	5.38	96.98	2.90	16.23	83.77	68.43	52.28	38.58
Quartile 4	5.37	96.49	2.96	14.71	85.29	69.44	55.07	39.79
$C \succ_i^F D$ and $\max(D) P_i \max(C)$								
All	2.02	80.74	0.54	66.10	33.90	14.46	4.92	0.78
Quartile 1	1.69	70.84	0.61	63.12	36.88	16.17	6.66	0.89
Quartile 2	2.07	81.65	0.50	68.55	31.45	13.02	4.50	0.74
Quartile 3	2.24	86.87	0.54	65.74	34.26	14.36	4.16	0.90
Quartile 4	2.10	83.66	0.53	66.99	33.01	14.30	4.33	0.57

Notes: The first column shows the mean number of responses per subject (in a total of 15 set comparisons) such that C (either set A or B) is considered to have more freedom than D (A if $C = B$ or B if $C = A$) and $\max(D) R_i \max(C)$ (first three rows), $\max(D) I_i \max(C)$ (second three rows), or $\max(D) P_i \max(C)$ (third three rows). The second column shows the percentage of participants for which the number of relevant cases is at least one. The third column shows the mean number of responses per subject such that $C \succ_i^W D$ in the relevant cases (i.e., when $C \succ_i^F D$ and $\max(D) R_i \max(C)$/$\max(D) I_i \max(C)$/$\max(D) P_i \max(C)$). The last five columns present the percentage of subjects who reveal IvFoC among the participants for which the number of relevant cases is at least one.

Table E.3. IvFoC and total response time

likely to reveal IvFoC than participants in *FreedomIU*: participants in *IUFreedom* exhibit IvFoC in 52% of relevant cases, while participants in *FreedomIU* exhibit it in 44% of these cases. This difference is statistically significant (see Table E.6). At the same time, overall results for $\max(B) R_i \max(A)$, $\max(B) I_i \max(A)$, and $\max(B) P_i \max(A)$ are again similar for both versions.

	IvFoC ($C \succ_i^W D$ in relevant cases)		
	$C \succ_i^F D$ and $\max(D) R_i \max(C)$	$C \succ_i^F D$ and $\max(D) I_i \max(C)$	$C \succ_i^F D$ and $\max(D) P_i \max(C)$
Quartile 1 (*Q1*)	(ref)	(ref)	(ref)
Quartile 2 (*Q2*)	-0.456***	-0.221***	-0.236***
	(0.0911)	(0.0707)	(0.0337)
Quartile 3 (*Q3*)	-0.595***	-0.339***	-0.257***
	(0.0936)	(0.0724)	(0.0346)
Quartile 4 (*Q4*)	-0.534***	-0.292***	-0.258***
	(0.0952)	(0.0738)	(0.0352)
Wald tests (*p*-value)			
$Q2 = Q3$	0.1258	0.0960	0.5333
$Q2 = Q4$	0.3976	0.3210	0.5230
$Q3 = Q4$	0.4987	0.5091	0.9804

Notes: This table reports the average marginal effects of the quartile for response time using OLS regression for the three measures of IvFoC, controlling for the number of "relevant cases", country, and observed individual characteristics.

Table E.4. IvFoC and total response time: Regression results

	Relevant cases		IvFoC ($C \succ_i^W D$ in relevant cases)					
	Mean (out of 15)	% subjects > 0	Mean	% of subjects s.t. IvFoC revelations $= 0$	≥ 1	≥ 2	≥ 3	≥ 4
$C \succ_i^F D$ and $\max(D) R_i \max(C)$								
All	7.15	96.43	3.42	15.12	84.88	70.60	57.67	45.84
FreedomIU	7.50	97.36	3.29	15.19	84.81	70.33	57.66	45.06
IUFreedom	6.83	95.54	3.54	15.05	84.95	70.86	57.68	46.58
$C \succ_i^F D$ and $\max(D) I_i \max(C)$								
All	5.13	95.19	2.87	17.50	82.50	66.67	51.96	38.43
FreedomIU	5.37	96.44	2.74	17.11	82.89	66.49	51.17	36.69
IUFreedom	4.90	93.99	3.00	17.87	82.13	66.84	52.71	40.09
$C \succ_i^F D$ and $\max(D) P_i \max(C)$								
All	2.02	80.74	0.54	66.10	33.90	14.46	4.92	0.78
FreedomIU	2.12	83.14	0.55	66.95	33.05	15.02	5.61	0.84
IUFreedom	1.93	78.46	0.54	65.29	34.71	13.93	4.26	0.72

Notes: The first column shows the mean number of responses per subject (in a total of 15 set comparisons) such that C (either set A or B) is considered to have more freedom than D (A if $C = B$ or B if $C = A$) and $\max(D) R_i \max(C)$ (first three rows), $\max(D) I_i \max(C)$ (second three rows), or $\max(D) P_i \max(C)$ (third three rows). The second column shows the percentage of participants for which the number of relevant cases is at least one. The third column shows the mean number of responses per subject such that $C \succ_i^W D$ in the relevant cases (i.e., when $C \succ_i^F D$ and $\max(D) R_i \max(C)$/$\max(D) I_i \max(C)$/$\max(D) P_i \max(C)$). The last five columns present the percentage of subjects who reveal IvFoC among the participants for which the number of relevant cases is at least one.

Table E.5. IvFoC and version of the survey

	IvFoC ($C \succ_i^W D$ in relevant cases)		
	$C \succ_i^F D$ and $\max(D) R_i \max(C)$	$C \succ_i^F D$ and $\max(D) I_i \max(C)$	$C \succ_i^F D$ and $\max(D) P_i \max(C)$
FreedomIU	(ref)	(ref)	(ref)
IUFreedom	0.545***	0.540***	0.0419*
	(0.0637)	(0.0492)	(0.0237)

Notes: This table reports the average marginal effects of the quartile for response time using OLS regression for the three measures of IvFoC, controlling for the number of “relevant cases”, country, and observed individual characteristics.

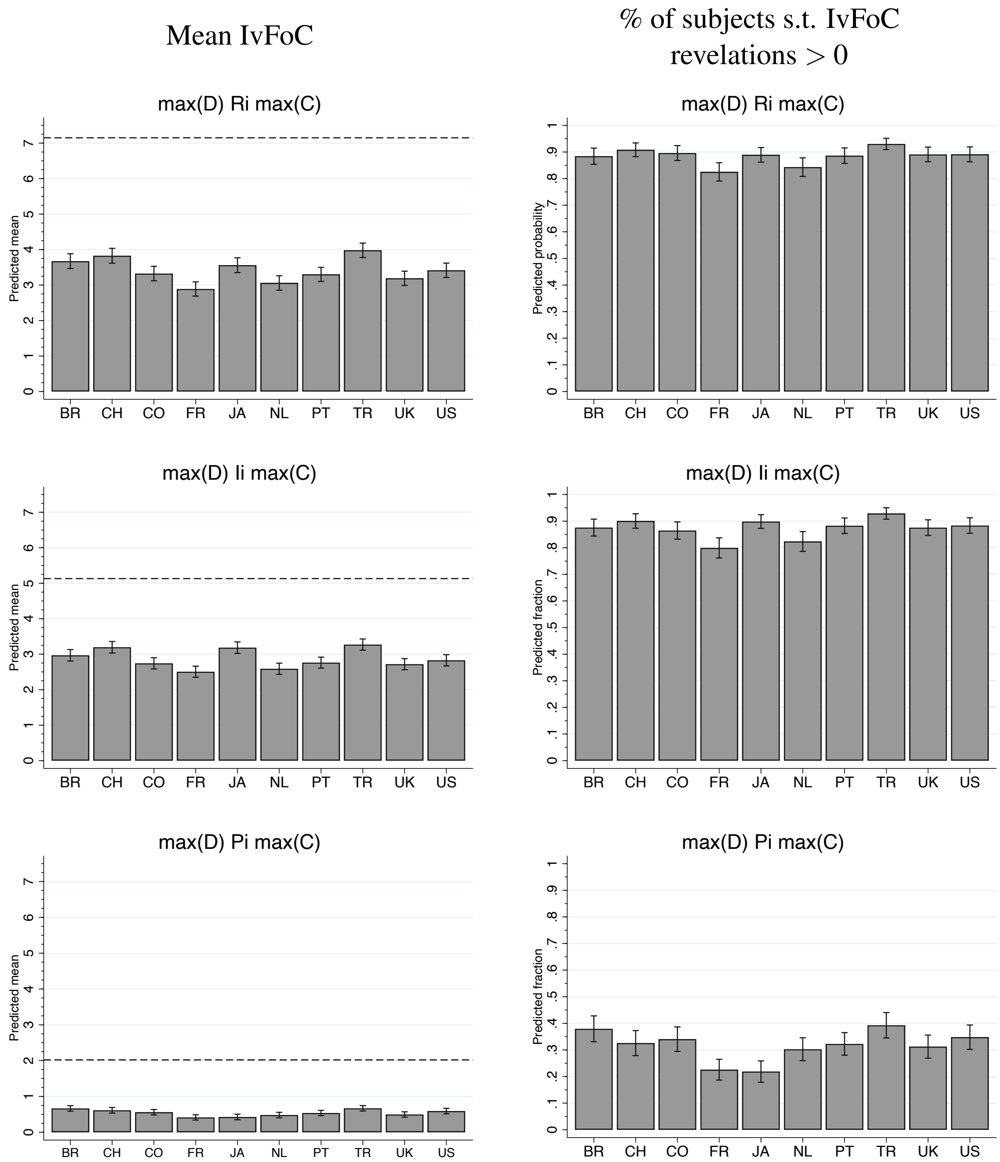
Table E.6. IvFoc and version of the survey: Regression results

F. ADDITIONAL RESULTS: CROSS-COUNTRY ANALYSIS

	Freedom		Welfare	
Cardinality family	(base)		(base)	
IU family				
BR	0.0913	(0.233)	0.0413	(0.187)
CH	0.541**	(0.230)	0.0318	(0.198)
CO	0.575***	(0.220)	0.248	(0.188)
FR	0.476**	(0.212)	0.607***	(0.206)
JA	0.638***	(0.216)	0.987***	(0.223)
NL	-0.142	(0.225)	0.318	(0.194)
PT	-0.131	(0.236)	0.170	(0.191)
TR	0.740***	(0.216)	-0.205	(0.184)
UK	-0.188	(0.226)	0.221	(0.193)
US	(ref)		(ref)	
Potential pref. family				
BR	0.0901	(0.381)	-0.146	(0.235)
CH	0.529	(0.359)	-0.0215	(0.242)
CO	0.443	(0.367)	0.196	(0.231)
FR	0.248	(0.355)	0.832***	(0.240)
JA	0.141	(0.390)	0.813***	(0.262)
NL	-0.0967	(0.380)	0.391*	(0.233)
PT	-0.711	(0.446)	0.435*	(0.226)
TR	0.396	(0.376)	-0.264	(0.229)
UK	-0.0220	(0.347)	0.552**	(0.227)
US	(ref)		(ref)	
Other rules				
BR	0.0400	(0.181)	0.0350	(0.195)
CH	0.705***	(0.177)	0.490**	(0.200)
CO	-0.0143	(0.187)	0.0358	(0.199)
FR	0.499***	(0.169)	0.731***	(0.213)
JA	0.743***	(0.170)	1.100***	(0.229)
NL	-0.352*	(0.183)	0.295	(0.202)
PT	-0.789***	(0.212)	0.182	(0.199)
TR	0.378**	(0.179)	-0.206	(0.193)
UK	-0.155	(0.180)	0.375*	(0.200)
US	(ref)		(ref)	
Observed characteristics	Yes		Yes	
Observations	4902		4902	
Log Lik	-4347.949		-6346.675	

Notes: This table reports multinomial regressions when the dependent variables are the family of rules assigned to the participants: Cardinality, Indirect utility, Potential preferences, and Other rules. Standard errors in parentheses. "Observed characteristics" include gender, age, social status, education, occupation and health status.

Table F.1. Classification: Multinomial regression results



Notes: The left panel reports the predicted average value of the number of responses such that C (either set A or B) is considered to provide more welfare than D (A if $C = B$ or B if $C = A$), when C is considered to have more freedom than D and $\max(D)R_i \max(C)$/$\max(D)I_i \max(C)$/$\max(D)P_i \max(C)$, using an OLS regression controlling for the number of relevant cases and observed characteristics. The right panel reports the predicted probability of these numbers being positive, using an OLS regression controlling for the same variables. See Table F.6 for underlying regressions. Lines indicate 95% confidence interval.

Figure F.7. Intrinsic value of freedom per country: All, indifference, and strict preference

Probability of a participant being assigned to a rule in:

Cardinality family

	BR	CH	CO	FR	JA	NL	PT	TR	UK
CH	0.000								
CO	0.220	0.007							
FR	0.006	0.220	0.121						
JA	0.000	0.869	0.004	0.154					
NL	0.035	0.000	0.001	0.000	0.000				
PT	0.000	0.000	0.000	0.000	0.000	0.083			
TR	0.003	0.337	0.067	0.776	0.269	0.000	0.000		
UK	0.168	0.000	0.009	0.000	0.000	0.466	0.015	0.000	
US	0.684	0.000	0.101	0.001	0.000	0.079	0.001	0.001	0.301

IU family

	BR	CH	CO	FR	JA	NL	PT	TR	UK
CH	0.234								
CO	0.028	0.345							
FR	0.242	0.946	0.287						
JA	0.118	0.727	0.534	0.651					
NL	0.532	0.067	0.005	0.052	0.020				
PT	0.718	0.119	0.011	0.113	0.046	0.807			
TR	0.011	0.169	0.691	0.131	0.303	0.001	0.003		
UK	0.317	0.025	0.001	0.018	0.006	0.682	0.525	0.000	
US	0.732	0.116	0.010	0.102	0.042	0.767	0.972	0.003	0.476

Potential preferences family

	BR	CH	CO	FR	JA	NL	PT	TR	UK
CH	0.485								
CO	0.404	0.885							
FR	0.977	0.481	0.411						
JA	0.660	0.249	0.212	0.617					
NL	0.822	0.362	0.306	0.788	0.823				
PT	0.131	0.029	0.023	0.109	0.268	0.194			
TR	0.669	0.794	0.696	0.673	0.386	0.513	0.056		
UK	0.898	0.377	0.318	0.863	0.724	0.911	0.140	0.561	
US	0.856	0.361	0.302	0.821	0.774	0.957	0.164	0.532	0.952

Notes: Table shows p-values for Wald tests of equality between countries based on regression estimates from Logit regressions, controlling for country and observed characteristics.

Table F.2. Classification for Freedom question: Statistical tests for country differences

Probability of a participant being assigned to a rule in:

Cardinality family

	BR	CH	CO	FR	JA	NL	PT	TR	UK
CH	0.259								
CO	0.351	0.815							
FR	0.000	0.012	0.005						
JA	0.000	0.000	0.000	0.178					
NL	0.074	0.519	0.372	0.058	0.002				
PT	0.194	0.880	0.691	0.015	0.000	0.605			
TR	0.175	0.011	0.020	0.000	0.000	0.002	0.007		
UK	0.060	0.446	0.311	0.073	0.002	0.910	0.524	0.001	
US	0.976	0.252	0.344	0.000	0.000	0.068	0.185	0.186	0.051

IU family

	BR	CH	CO	FR	JA	NL	PT	TR	UK
CH	0.071								
CO	0.359	0.007							
FR	0.817	0.108	0.249						
JA	0.317	0.005	0.902	0.199					
NL	0.849	0.047	0.473	0.661	0.397				
PT	0.546	0.210	0.121	0.702	0.096	0.421			
TR	0.421	0.297	0.082	0.556	0.069	0.310	0.831		
UK	0.259	0.480	0.040	0.346	0.027	0.175	0.583	0.745	
US	0.704	0.149	0.196	0.882	0.156	0.562	0.820	0.664	0.432

Potential preferences family

	BR	CH	CO	FR	JA	NL	PT	TR	UK
CH	0.479								
CO	0.046	0.191							
FR	0.019	0.086	0.708						
JA	0.215	0.566	0.470	0.258					
NL	0.098	0.338	0.731	0.454	0.694				
PT	0.009	0.048	0.539	0.812	0.170	0.333			
TR	0.486	0.989	0.181	0.080	0.559	0.319	0.046		
UK	0.008	0.042	0.486	0.730	0.142	0.284	0.917	0.041	
US	0.089	0.313	0.763	0.484	0.662	0.964	0.357	0.303	0.298

Notes: Table shows p-values for Wald tests of equality between countries based on regression estimates from Logit regressions, controlling for country and observed characteristics.

Table F.3. Classification for Welfare question: Statistical tests for country differences

Average number of responses such that $C \succ_i^W D$ when:

$C \succ_i^F D$ and $\max(D) R_i \max(C)$

	BR	CH	CO	FR	JA	NL	PT	TR	UK
CH	0.314								
CO	0.018	0.001							
FR	0.000	0.000	0.003						
JA	0.461	0.081	0.120	0.000					
NL	0.000	0.000	0.075	0.241	0.001				
PT	0.012	0.000	0.875	0.004	0.079	0.098			
TR	0.040	0.290	0.000	0.000	0.005	0.000	0.000		
UK	0.001	0.000	0.364	0.036	0.012	0.369	0.441	0.000	
US	0.083	0.007	0.536	0.000	0.336	0.015	0.435	0.000	0.117

$C \succ_i^F D$ and $\max(D) I_i \max(C)$

	BR	CH	CO	FR	JA	NL	PT	TR	UK
CH	0.053								
CO	0.046	0.000							
FR	0.000	0.000	0.041						
JA	0.071	0.926	0.000	0.000					
NL	0.001	0.000	0.191	0.455	0.000				
PT	0.074	0.000	0.843	0.021	0.000	0.127			
TR	0.009	0.511	0.000	0.000	0.465	0.000	0.000		
UK	0.032	0.000	0.847	0.056	0.000	0.255	0.692	0.000	
US	0.220	0.002	0.452	0.004	0.002	0.036	0.573	0.000	0.332

$C \succ_i^F D$ and $\max(D) P_i \max(C)$

	BR	CH	CO	FR	JA	NL	PT	TR	UK
CH	0.360								
CO	0.052	0.325							
FR	0.000	0.000	0.008						
JA	0.000	0.001	0.018	0.828					
NL	0.001	0.018	0.155	0.221	0.331				
PT	0.019	0.156	0.673	0.022	0.044	0.302			
TR	0.947	0.312	0.044	0.000	0.000	0.001	0.014		
UK	0.002	0.034	0.252	0.125	0.201	0.769	0.454	0.002	
US	0.179	0.687	0.559	0.001	0.003	0.041	0.315	0.160	0.076

Notes: Table shows p-values for Wald tests of equality between countries based on regression estimates reported in columns 1, 3, and 5 of Table F.6.

Table F.4. IvFoC (mean): Statistical tests for country differences

Probability that at least one response per subject reveals $C \succ_i^W D$ when:

$C \succ_i^F D$ and $\max(D) R_i \max(C)$

	BR	CH	CO	FR	JA	NL	PT	TR	UK
CH	0.232								
CO	0.566	0.526							
FR	0.014	0.000	0.002						
JA	0.810	0.318	0.731	0.003					
NL	0.081	0.004	0.021	0.466	0.037				
PT	0.923	0.263	0.632	0.008	0.883	0.064			
TR	0.013	0.181	0.050	0.000	0.019	0.000	0.015		
UK	0.749	0.365	0.796	0.003	0.931	0.033	0.820	0.026	
US	0.740	0.378	0.807	0.003	0.922	0.033	0.814	0.027	0.990

$C \succ_i^F D$ and $\max(D) I_i \max(C)$

	BR	CH	CO	FR	JA	NL	PT	TR	UK
CH	0.239								
CO	0.625	0.094							
FR	0.003	0.000	0.011						
JA	0.276	0.911	0.106	0.000					
NL	0.035	0.001	0.105	0.363	0.001				
PT	0.752	0.373	0.411	0.001	0.417	0.014			
TR	0.005	0.096	0.001	0.000	0.073	0.000	0.011		
UK	0.990	0.218	0.626	0.001	0.237	0.030	0.730	0.004	
US	0.725	0.397	0.397	0.000	0.439	0.011	0.974	0.012	0.703

$C \succ_i^F D$ and $\max(D) P_i \max(C)$

	BR	CH	CO	FR	JA	NL	PT	TR	UK
CH	0.120								
CO	0.245	0.662							
FR	0.000	0.001	0.000						
JA	0.000	0.001	0.000	0.806					
NL	0.020	0.484	0.245	0.009	0.006				
PT	0.086	0.923	0.574	0.001	0.001	0.518			
TR	0.703	0.047	0.121	0.000	0.000	0.006	0.031		
UK	0.046	0.683	0.387	0.003	0.002	0.754	0.738	0.017	
US	0.353	0.512	0.822	0.000	0.000	0.155	0.430	0.189	0.260

Notes: Table shows p-values for Wald tests of equality between countries based on regression estimates reported in columns 2, 4, and 6 of Table F.6.

Table F.5. IvFoC (revelations > 0): Statistical tests for country differences

	IvFoC ($C \succ_i^W D$ in relevant cases)					
	$C \succ_i^F D$		$C \succ_i^F D$		$C \succ_i^F D$	
	$\max(B) R_i \max(A)$		$\max(B) I_i \max(A)$		$\max(B) P_i \max(A)$	
	(1)	(2)	(3)	(4)	(5)	(6)
	Mean	Revelations > 0	Mean	Revelations > 0	Mean	Revelations > 0
BR	0.257*	-0.0697	0.141	-0.0719	0.0739	0.136
	(0.148)	(0.210)	(0.115)	(0.205)	(0.0550)	(0.147)
CH	0.409***	0.190	0.368***	0.178	0.0226	-0.0994
	(0.151)	(0.215)	(0.117)	(0.210)	(0.0560)	(0.152)
CO	-0.0915	0.0516	-0.0865	-0.170	-0.0321	-0.0332
	(0.148)	(0.211)	(0.115)	(0.201)	(0.0550)	(0.147)
FR	-0.526***	-0.550***	-0.322***	-0.643***	-0.177***	-0.605***
	(0.145)	(0.187)	(0.113)	(0.184)	(0.0539)	(0.153)
JA	0.143	-0.0196	0.357***	0.155	-0.165***	-0.645***
	(0.149)	(0.201)	(0.116)	(0.200)	(0.0553)	(0.157)
NL	-0.356**	-0.421**	-0.238**	-0.485**	-0.111**	-0.206
	(0.146)	(0.197)	(0.113)	(0.192)	(0.0542)	(0.145)
PT	-0.114	-0.0491	-0.0642	-0.00659	-0.0547	-0.113
	(0.147)	(0.209)	(0.114)	(0.203)	(0.0544)	(0.144)
TR	0.564***	0.488**	0.443***	0.542**	0.0775	0.193
	(0.148)	(0.221)	(0.115)	(0.217)	(0.0551)	(0.147)
UK	-0.225	-0.00259	-0.109	-0.0744	-0.0948*	-0.161
	(0.144)	(0.202)	(0.112)	(0.195)	(0.0534)	(0.143)
US	(ref)	(ref)	(ref)	(ref)	(ref)	(ref)
# of relevant cases	0.455***	0.337***	0.608***	0.523***	0.273***	0.609***
	(0.0115)	(0.0147)	(0.0114)	(0.0199)	(0.00924)	(0.0290
Observed characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4,902	4,902	4,902	4,902	4,902	4,902
R-squared	0.268		0.387		0.172	
Pseudo R-squared		0.163		0.207		0.103
Log Lik		-1743		-1803		-2815

Notes: Columns 1, 3, and 5 based on OLS regressions, and columns 2, 4, and 6 based on Logit regressions. Standard-errors in parentheses. “Relevant cases” means set comparisons for which $C \succ_i^F D$ and $\max(D) R_i \max(C)$/$\max(D) I_i \max(C)$/$\max(D) P_i \max(C)$. “Observed characteristics” include gender, age, social status, education, occupation and health status.

Table F.6. Intrinsic value of freedom: Regression with country dummies

G. INSTRUCTIONS

This appendix transcribes the instructions, questions, and choice options given to participants in their screens.[29] For concision, we only present one set comparison and the attention check ("Screen 5" and "Attention check" below), while subjects were presented with 15 set comparisons and one attention check, all in a random order. The instructions for all set comparisons are equivalent with the exception of the hospitals available. The instructions for the set comparison and main questions are of the *FreedomIU* version. The *IUFreedom* version is equivalent except for the order of the Freedom and IU questions. Comments that are not part of the instructions are shown in square brackets.

[Screen 1]

What is your gender?

[] Male

[] Female

[Screen 2]

How old are you?

[text entry, answer in the range 0 to 99]

[Screen 3]

Dear participant, thank you very much for joining this survey!

This survey is organized by a consortium of universities (University of Lyon (France), University of Osaka (Japan), University of Southampton (United Kingdom), and its results will be used in an academic research project.

Most people will take around 15 minutes to complete this questionnaire. **All answers are anonymous:** that means that your answers cannot be traced back to you.

During the survey, you are going to be asked several questions concerning hypothetical situations. Please reflect upon the questions before answering. **There are no right or wrong answers, we are just interested in your views.**

[29] Translations for other languages used in our research are available upon request.

Your input is valuable for our research project. Thank you!

[Screen 4]

Imagine that two individuals, **Mr. Green and Mr. Yellow**, are similar in all respects. They have to undergo a surgical procedure that is of minimal risk to their overall health. However, this surgery means that they will need to spend four days recovering in hospital.

The two individuals each have to choose a hospital for this surgery and for the recovery time. However, the hospitals available to each individual depends on each individual's health insurance plan.

The hospitals available are equivalent in terms of surgery care quality, doctors' skills, etc. Thus, Mr. Green and Mr. Yellow's overall health will not be affected by the choice of hospital for the surgery and the recovering stay.

However, the hospitals differ in terms of general staff quality (excluding doctors) and general comfort quality. A trustworthy non-government agency that rates hospitals in their country provides the following information:

- The **staff quality** (excluding doctors) relates to service & assistance quality, nursing quality, friendliness of staff, etc.
- The **comfort quality** relates to bed quality, food quality, amenities, etc.

Each dimension is rated from 1% for the lowest quality to 100% for the highest quality. Consider the following example:

- Hospital A: staff 78%, comfort 65%
- Hospital B: staff 62%, comfort 89%

In this example, Hospital A provides better quality of staff but a lower level of comfort than Hospital B.

In the following, you are going to be presented with several hypothetical situations. For all of them, you are asked to compare the situation of Mr. Green and of Mr. Yellow, which differ according to their health insurance plan.

[Screen 5]

Mr. Green has the following hospital available in his insurance plan:

- **Hospital A (staff 80%, comfort 80%)**

Mr. Yellow has the following hospital available in his insurance plan:

- **Hospital B (staff 70%, comfort 90%)**

Which individual do you think has more freedom of choice?

[] Mr. Green

[] Mr. Yellow

[] The same

Which hospital do you think is the best for the treatment and recovery time?

You can select more than one hospital if you think two or more hospitals are equally best

[] Hospital A

[] Hospital B

All things considered, which individual do you think has the best insurance plan?

[] Mr. Green

[] Mr. Yellow

[] Equally good

[Attention check]

Mr. Green has the following hospital available in his insurance plan:

- **Hospital C (staff 80%, comfort 80%)**

Mr. Yellow has the following hospital available in his insurance plan:

- **Hospital C (staff 80%, comfort 80%)**

Which individual do you think has more freedom of choice?

[] Mr. Green

[] Mr. Yellow

[] The same

Which hospital do you think is the best for the treatment and recovery time?

You can select more than one hospital if you think two or more hospitals are equally best

[] Hospital C

[] Hospital C

All things considered, which individual do you think has the best insurance plan?

[] Mr. Green

[] Mr. Yellow

[] If you are not a robot please click on this button

[Screen 21]

Nearly done! Please click to pass to the final small set of questions.

[Screen 22]

On a scale of 1 (very difficult) to 10 (very easy), how difficult was it for you to understand the survey?

[1 to 10 scale]

[Screen 23]

In general, would you say that your health is:

[] Excellent

[] Very good

[] Good

[] Fair

[] Poor

[Screen 24]

People sometimes describe themselves as being working class, middle class, upper class or lower class. Would you describe yourself as being:

[] Upper class

[] Upper middle class

[] Lower middle class

[] Working class

[] Lower class

[Screen 25]

What is the highest level of education that you have completed?

[] Early childhood education

[] Primary education

[] Lower secondary education

[] Upper secondary education

[] Post-secondary non-tertiary education

[] Bachelor's or equivalent level

[] Master's or equivalent level

[] Doctoral or equivalent level

[Screen 26]

Are you currently:

[] Employed

[] Self-employed

[] Unemployed

[] Retired

[] Student

[] Inactive