Latent Relevance and the Identification of Deliberate Preference Change*

Niels Boissonnet†

Abstract

A decision maker changes her preferences deliberately when she chooses which values to treat as relevant by maximizing a stable metapreference. Choices, however, identify relevant values only partially: a value can be relevant at a period without affecting any choice then, in which case it is latent. For data that can be rationalized with no value latent, we characterize those that leave no room for latent values: at every period, every value must be revealed relevant by that period's choices, or revealed irrelevant by the requirement that a single way of weighing values fit every period. When the data leave no room for latent values, which is the case in particular when every value makes a difference on its own at some period, testing deliberation as if none were latent is without loss. Otherwise, latent values can turn an apparent reversal into a deliberate change. In a population, individuals can share a metapreference exactly when some relevance paths that fit their choices are each deliberate and form no cycle of conflicting moves, and two observations per individual can suffice to reject such a shared view.

Keywords: Preference change · Revealed preference · Partial identification · Values · Metapreferences

JEL Classification: D01 · D11 · D91

*This paper comes in part from joint work with Alexis Ghersengorin, with whom I developed the initial model on which it builds and first stated the problem it studies. I am deeply grateful to him for countless discussions and for his contributions to that work. I also thank Simon Gleyze for discussions on the initial model, and Gerrit Bauch, Yves Breitmoser, Franz Dietrich, Frank Riedel, participants in the economic theory seminar at Bielefeld and seminar participants at Paris 1 Panthéon-Sorbonne for helpful comments. Financial support from the Ministry of Culture and Science of North Rhine-Westphalia is gratefully acknowledged.

†Center for Mathematical Economics, Bielefeld University, Bielefeld, Germany. Email: niels.boissonnet@gmail.com. ORCID: 0009-0006-3639-3098.

1 Introduction

Consider two recruiters whose choices are, at a given moment, exactly the same: both believe in meritocracy, and neither is swayed by the gender of applicants. Yet they differ in a way their choices do not show. The first regards gender as a legitimate consideration; the second does not. The difference shows once an experience leads them to reconsider their values. Suppose that both read Sandel (2020), who argues that success owes more to luck and background than the successful like to think. The argument speaks to the first recruiter: if starting points are unequal, gender, which she already regards as legitimate, should count. She may come to see meritocracy as a myth and become open to affirmative action. The second is harder to persuade. For her, gender is beside the point, and the argument has less to build on: she may grant that luck plays its part and still take merit to be the fairest guide she has. For the first recruiter, the relevance of gender is *latent*: it plays no role in her present choices, yet it can shape how her preferences evolve.

What makes latent values intriguing is that they are hard to infer from past choices, yet they can drive social change. In the mid-eighteenth century, reformers such as John Woolman persuaded Quakers to give up slavery mainly by appealing to principles they already professed; in 1776 the Philadelphia Yearly Meeting prohibited slaveholding among its members (Soderlund, 1985).[1] For these Quakers, the testimony of equality was *latent*: it played no role in their decision to own slaves, yet it gave the reformers their lever and eventually changed their choices. Social psychology finds the same pattern in experiments. Conservatives rarely view environmental issues in moral terms, yet they become markedly more supportive of environmental protection when the argument appeals to purity, a value they hold but do not ordinarily bring to bear on the environment (Feinberg and Willer, 2013). Similarly, a ten-minute conversation inviting voters to understand the experience of transgender people by likening it to an experience of their own reduced prejudice for at least three months (Broockman and Kalla, 2016). In each case, what changed

[1] Part of this shift came from the deaths of influential slaveholders, not only from individuals revising their views.

people's minds was a consideration they already held but did not use.

This paper investigates how preference change can be modeled when some values are latent. A natural way to model preference change is to trace it to a modification of the values that underpin the decision maker's preferences (Dietrich and List, 2013a, 2016; Bernheim et al., 2021). We adopt this view through a variant of the model of Boissonnet et al. (2023). Each alternative is described by the values it has. At each period, the decision maker (DM) deems some values relevant—her *relevance set*—and ranks alternatives by comparing them on these values according to a *weighing relation* that is the same at every period. Change is deliberate when, after reflecting on some values, whether or not they currently drive her behavior, DM chooses which of them to take up or set aside by maximizing a stable *metapreference* over the relevance sets available to her. Whether a change is deliberate thus depends on what DM deemed relevant before and after it.

Our contribution is twofold. First, among data that can be rationalized with no value latent, we characterize those that leave no room for latent values (Theorem 1). A value that affects choices at some period can be latent at another only if it makes a difference solely in combination with values DM does not then deem relevant, as gender does with the field for the first recruiter. When every value makes a difference on its own at some period, nothing can be latent, and testing deliberation as if no value were latent is without loss. Otherwise, as for the first recruiter, an apparent reversal may be a deliberate change. A test has full power only if every value the analyst lists affects some choice: values that never do can be switched on and off at will, and enough of them make deliberate any sequence of preferences that a single weighing relation fits.

Second, we turn to populations. Many panels observe each person only twice, before and after a campaign or an intervention,[2] and with two observations any change of a single person that one weighing relation fits can be read as deliberate. Comparing people does better. We ask when individuals whose preferences move

[2]Such panels record far less than our data, which consist of a complete ranking of all combinations of values at each observation. Our results say what such data can identify in principle; they are not a ready-made test for existing panels.

apart can share a view of which values should matter, that is, a common metapreference. Divergence alone does not contradict such a view: individuals who reconsider different values can end up far apart. What contradicts it is a crossing: two individuals who have each had available the relevance set the other now holds, and settled on another—or, with more individuals, a cycle of such moves (Theorem 2). A common metapreference can thus be rejected with two observations per individual, and for two individuals observed twice we characterize the data that reject it (Proposition 2).

Section 2 presents the model and its two characterizations, Section 3 studies latent values, and Section 4 turns to populations. Appendix A collects the proofs, and Appendix B treats transitive weighing relations.

1.1 Related literature

Our model is a variant of that of Boissonnet et al. (2023), who test deliberation assuming that no value is latent; for binary attributes, their characterization is the special case of ours in which DM's relevance sets are those her choices reveal. An earlier version of their paper notes that relevance is not identified in general: under richness and transitivity, the relevance vectors that fit the data form a lattice whose minimum is the revealed one (Boissonnet et al., 2022, Proposition 4). It removes this indeterminacy by assuming that indifference comes only from irrelevance; we study it instead. They also address a related concern, that an analyst could add values until the model fits, by having the analyst commit to a list of values in advance. Our problem arises after this commitment: it concerns which of the listed values are relevant, not which values exist.

This paper comes in part from unpublished joint work with Alexis Ghersengorin, which states this problem and calls latent values background attributes. For a transitive weighing relation, that work characterizes deliberation through the changes an analyst attributes to DM between periods. Corollary 1 is the counterpart of this result for relevance paths, without transitivity, and the Deliberate condition is its acyclicity condition written on relevance sets. As that work observes, its conditions

are not conditions on choices. Here, we ask which values the data allow to be latent, and we extend the analysis to populations.

The weighing relation is borrowed from the reason-based theory of Dietrich and List (2013b,a, 2016), in which preferences derive from sets of motivationally salient properties and change as salience changes. Closest to our question, Dietrich and List (2013c) count a property as motivationally salient if and only if it makes a behavioral difference, which selects one model among those that fit a choice function. Revealed relevance plays the same role here, and we ask what such a selection leaves out: a value can be relevant without making a difference, and this matters once preferences change. Bernheim et al. (2021) model the deliberate choice of worldviews, Grüne-Yanoff and Hansson (2009) survey approaches to preference change, and the metapreference connects to the philosophical literature on higher-order preferences (Frankfurt, 1971; Jeffrey, 1974). Pivato (2025a,b) formalizes such preferences over preferences and relates them to autonomy. Hayashi (2026) represents layers of metapreferences recursively, with preference change coming from the current self's judgment about its successor. Ertemel (2026) studies deterministic choices before and after a deliberative intervention on two fixed dimensions, and also finds that an abandoned alternative does not return. In models of salience and focusing (Bordalo et al., 2013; Kőszegi and Szeidl, 2013), what the decision maker attends to is driven by the choice context rather than chosen. These approaches ask how preferences derive from what the decision maker attends to. We ask what choices reveal about it, a question that applies to them as well.

Methodologically, the paper belongs to revealed preference theory (Richter, 1966). Relevance is only partially identified, as are consideration sets in models of limited attention (Masatlioglu et al., 2012). In populations, our question is related to tests of how much preference heterogeneity a population requires (Crawford and Pendakur, 2013); here, individuals may differ in how they weigh values, and we test whether they can share a ranking of relevance sets. Value-based evaluation goes back to Lancaster (1966) and Tversky (1972).

2 Model

2.1 Deliberate preference change

Let $\mathcal{M}$ be a finite set of values, with $|\mathcal{M}| \geq 2$.[3] Each value is either present or absent, so alternatives are subsets of $\mathcal{M}$, and they lie in $X = 2^{\mathcal{M}}$.[4] In the recruiting example, values may be holding a degree, being a woman, and being trained in a field where women are underrepresented, and a candidate is described by the values she has. So values are treated like attributes.

The model rests on two ideas (Dietrich and List, 2013a; Boissonnet et al., 2023). First, preferences derive from the values DM deems relevant. At a given time, DM deems some set $M \subseteq \mathcal{M}$ relevant, her *relevance set*, and evaluates an alternative x only through $x \cap M$. She compares such combinations with a *weighing relation*, a complete binary relation $\succcurlyeq$ on $\Omega = 2^{\mathcal{M}}$, with asymmetric part $\succ$ and symmetric part $\simeq$. A pair $(M, \succcurlyeq)$ *rationalizes* a preference $\succsim$ on X if

$$x \succsim y \iff x \cap M \succcurlyeq y \cap M \qquad \text{for all } x, y \in X.$$

The weighing relation does not change over time, so preferences change only because the relevance set does. We do not require it to be transitive. Whether a transitive weighing relation fits the data depends only on how DM ranks combinations of the values she reveals at each period, not on which values are latent (Appendix B). In many cases one does, in particular whenever DM is observed only twice and some weighing relation fits.

Second, DM can change her relevance set deliberately. Between two periods, she reconsiders the relevance of some values—those she becomes aware of, say, or those an intervention draws her attention to—and may switch any of them on or off. This is therefore a theory of deliberate preference change: between two periods, DM has the opportunity to revise some of the values that underpin her behavior. Given her

[3] The analyst fixes this list in advance, as in Boissonnet et al. (2023). Some values on it may never affect DM's choices; footnote 8 discusses them.

[4] Our concern is the identification problem due to latent values, not to missing options. Hence, we maintain the strong assumption that every combination of values is an available alternative.

current relevance set M and a set $A \subseteq \mathcal{M}$ of values she reconsiders, the relevance sets she can reach are

$$R(M, A) = \{M \triangle B : B \subseteq A\}.$$

Among them, she picks the best according to a *metapreference*, a linear order $\rhd$ on $2^{\mathcal{M}}$.[5] DM can thus question which values matter to her, but cannot rebuild her preferences from scratch.

Definition 1. A sequence $(M_t)_{t=1}^T$ of relevance sets is *generated by deliberation* if, for some metapreference $\rhd$ and some sets $(A_t)_{t<T}$ of reconsidered values, $M_{t+1} = \max(R(M_t, A_t), \rhd)$ for all $t < T$.

To see the kind of situation we have in mind, return to the two recruiters. Reading Sandel (2020) leads both to reconsider whether unequal starting points should count, for instance whether a candidate was trained in a field where women are underrepresented. The book thus fixes the values they reconsider; their metapreferences decide what they do with them. The first decides that the field of training matters. For her, gender matters only together with the field: she favors a woman trained in a field where women are underrepresented. So once she takes up the field, gender, which she already deemed relevant, starts to affect her choices. The second, for whom gender is not relevant, may grant the point and still keep her relevance set unchanged. The two can diverge even if they share a metapreference, because they start from different relevance sets.

2.2 Rationalization and relevance path

The analyst observes DM's preferences at $T \geq 2$ periods, complete preorders $(\succsim_t)_{t=1}^T$ on X, with asymmetric parts $\succ_t$ and symmetric parts $\sim_t$.

Definition 2. The data are *deliberate* if there are relevance sets $(M_t)_t$, generated by deliberation, and a weighing relation $\geqslant$ such that $(M_t, \geqslant)$ rationalizes $\succsim_t$ for every t.

[5] Taking a partial order and requiring the next relevance set to dominate every other reachable set is equivalent: such a partial order extends to a linear order with the same maxima (Szpilrajn, 1930).

The analyst does not see relevance sets. Choices only show which values make a difference.

Definition 3. A value m is *revealed relevant at period* t if $x \not\sim_t y$ for some $x, y \in X$ that differ only in m. Let U_t denote the set of values revealed relevant at t, and $\overline{M} = \bigcup_t U_t$.

For instance, gender is revealed relevant at a period exactly when the recruiter ranks differently two candidates who differ only in gender. Taken one period at a time, this is all the data say about relevance.

Lemma 1. *Let t be a period and $M \subseteq \mathcal{M}$. Some weighing relation $\geqslant$ is such that $(M, \geqslant)$ rationalizes $\succsim_t$ if and only if $U_t \subseteq M$.*

A value that affects a choice must be relevant. But a value that affects no choice may be relevant as well: taken one period at a time, the data cannot exclude that a value is relevant. Across periods they sometimes can, because the weighing relation is the same at every period.

A *relevance path* is a sequence $(M_t)_t$ with $U_t \subseteq M_t \subseteq \mathcal{M}$ for all t. It is *rationalizable* if a single weighing relation $\geqslant$ is such that $(M_t, \geqslant)$ rationalizes $\succsim_t$ for every t. A value of $M_t \setminus U_t$ is *latent* at t: it is relevant, but no choice at t reveals it. The smallest relevance path is $(U_t)_t$, along which no value is latent. Since the weighing relation is the same at every period, a path is rationalizable exactly when any two periods rank alike the combinations of values relevant at both (Lemma 3 in Appendix A). The more values are latent, the more combinations two periods share, and the more demanding this requirement.

2.3 Deliberation

When can relevance sets come from deliberation? Typically, the analyst does not observe which values DM reconsiders. Reconsidering more values only adds sets that DM must have turned down, so the most permissive guess is that she reconsiders exactly what she changes, $A_t = M_t \triangle M_{t+1}$. What deliberation then requires is the following.

Definition 4. A sequence $(M_t)_t$ of relevance sets is *Deliberate* if, for all t and all $\tau \geq t+2$,

$$M_t \triangle M_\tau \subseteq M_t \triangle M_{t+1} \implies M_\tau = M_{t+1}.$$

The Deliberate condition says that a relevance set, once turned down, is never chosen later. The premise says that M_τ differs from M_t only in values DM changed between t and $t+1$, so it was available to her at t. She turned it down in favor of M_{t+1}, so she cannot move away from M_{t+1} and settle on that set later.[6] Consider a recruiter who drops gender between the first and second periods and takes it up again between the second and third. At the third period, she settles on a set she had turned down at the first: the sequence is not Deliberate. If she also drops a value she had not reconsidered at the first period, the set she settles on was not available then, and no decision is reversed.

Proposition 1. *A sequence $(M_t)_{t=1}^T$ of relevance sets is generated by deliberation if and only if it is Deliberate.*

The Deliberate condition thus exhausts what deliberation implies. The proof also shows that nothing is lost by assuming that DM reconsiders exactly what she changes, $A_t = M_t \triangle M_{t+1}$.[7] Together with rationalizability, this gives the following.

Corollary 1. *The data are deliberate if and only if some rationalizable relevance path is Deliberate.*

The difficulty lies in the word "some". If one assumes that no value is latent, only $(U_t)_t$ is left, and Corollary 1 becomes a test on the data: the counterpart of the characterization of Boissonnet et al. (2023) for binary attributes. Otherwise, the analyst must search over relevance paths, and the data alone do not say which one is right. Section 3 studies this search.

[6] Applied to $(U_t)_t$, the Deliberate condition is the Deliberation axiom of Boissonnet et al. (2023), stated in contrapositive form.

[7] For $(U_t)_t$ and a partial metapreference, the result is the metapreference part of Theorem 1 of Boissonnet et al. (2023). Their argument never uses how relevance sets are identified, so the result holds for any relevance path. Our proof differs in one step: instead of the smallest partial order that ranks each M_t above the sets reachable before t, we build a linear order directly, ranking sets by the last period at which they appear. When the reconsidered values are observed, for instance because awareness is elicited or manipulated, deliberation holds if and only if each M_{t+1} belongs to $R(M_t, A_t)$ and the relation that places each M_{t+1} above every other set of $R(M_t, A_t)$ is acyclic.

3 Latent relevance

Why should the analyst care which relevance path is right? Because latent values can turn an apparent reversal into a deliberate change, as the following example shows.

Example 1. Return to the first recruiter of the introduction, and let her belief in merit take the form of caring about a degree. Candidates are described by three values: holding a degree (e), being a woman (g), and being trained in a field where women are underrepresented (f). Let $u(S) = 2 \cdot \mathbf{1}[e \in S] + \mathbf{1}[\{g, f\} \subseteq S]$. At periods 1 and 3, she ranks candidates by whether they hold a degree; at period 2, by u. Then $U_1 = U_3 = \{e\}$ and $U_2 = \{e, g, f\}$. If no value is latent, her relevance sets are $\{e\}$, $\{e, g, f\}$ and $\{e\}$. At period 3, she returns to a set she turned down at period 1: $U_1 \triangle U_3 = \emptyset$ is contained in $U_1 \triangle U_2 = \{g, f\}$, yet $U_3 \neq U_2$, so the path $(U_t)_t$ is not Deliberate.

Now suppose that she already regards gender as relevant at period 1, and consider the path $(M_1, M_2, M_3) = (\{e, g\}, \{e, g, f\}, \{e, f\})$. Since $u(x \cap \{e, g\}) = u(x \cap \{e, f\}) = 2 \cdot \mathbf{1}[e \in x]$ and $u(x \cap \{e, g, f\}) = u(x)$, the weighing relation represented by u rationalizes every period along this path; gender is latent at period 1, and the field at period 3. Now $M_1 \triangle M_3 = \{g, f\}$ is not contained in $M_1 \triangle M_2 = \{f\}$, so the path is Deliberate and, by Corollary 1, the data are deliberate. A metapreference with $\{e, f\} \rhd \{e, g, f\} \rhd \{e, g\}$ generates the path, with $A_1 = \{f\}$ and $A_2 = \{g\}$.

At period 1, the recruiter regards gender as relevant but cannot act on it, because the field of training, without which gender does not matter to her, is not yet among her concerns. When she takes up the field, gender starts to matter. When she later sets gender aside, she recovers her period-1 ranking without reversing any decision: the set she then holds required dropping gender, which she had not reconsidered at period 1. Assuming that no value is latent reads the same choices as a reversal.

Can the analyst then make values latent wherever it helps? Latent values never help to fit preferences: removing values from a rationalizable path keeps it rationalizable (Lemma 5 in Appendix A). What they offer is room for deliberation, as in

Example 1. But this room is limited, because a single weighing relation must still fit every period. Suppose that m is revealed at r but not at t, and make it relevant at t. The two periods must then rank alike all combinations of $(U_t \cap U_r) \cup \{m\}$ (Lemma 3). At t, adding m to such a combination changes nothing, since t does not reveal m. So it must change nothing at r either. If it does change something at r, then m cannot be latent at t. The following definition singles out the values for which such a conflict arises.

Definition 5. Let t be a period and $m \in \mathcal{M} \setminus U_t$. We say that m is *revealed irrelevant at* t if, for some period r with $m \in U_r$ and some $w \subseteq U_t \cap U_r$, $w \cup \{m\} \not\sim_r w$.

At some period that reveals m, adding m changes the ranking of a combination of values also revealed at t. Revealed relevance and revealed irrelevance have different sources. Revealed relevance bears on the choices of a single period: the choices at t alone show that m is relevant at t. Revealed irrelevance comes from the dynamic constraint of rationalization, that a single weighing relation fits every period: if m were relevant at t, periods t and r would have to rank alike combinations that they rank differently. Such a value is therefore irrelevant at t along every rationalizable path (Lemma 6 in Appendix A). Conversely, the proof of Theorem 1 shows that, if $(U_t)_t$ is rationalizable, every other value outside U_t is latent at t along some rationalizable path. A latent value also has a restricted effect. If $m \in M_t \setminus U_t$ along a rationalizable path and $m \in U_r$, then $w \cup \{m\} \sim_r w$ for every $w \subseteq M_t \cap M_r$: the two sets differ only in $m \notin U_t$, so they are indifferent at t, hence at r by Lemma 3. At r, m can make a difference only in combination with values outside M_t, which DM ignores at t.

Theorem 1. *Suppose that $(U_t)_t$ is rationalizable. It is the only rationalizable relevance path if and only if, at every period, every value is either revealed relevant or revealed irrelevant.*

The condition of Theorem 1 can fail in two ways. First, a value that no choice reveals is never revealed irrelevant, so it can be made relevant anywhere.[8] Second,

[8]This weakens every test of deliberation in this paper. Suppose that $\mathcal{M} \setminus \overline{M}$ contains $T-2$ distinct values $d_3, \ldots, d_T$, let D be their set, and take any rationalizable relevance path $(M_t)_t$. Let DM switch

a value that some choice reveals is not revealed irrelevant at t when it makes a difference only on combinations that involve values t does not reveal. A value that makes a difference on its own at some period s, that is, $\{m\} \not\sim_s \emptyset$, is therefore revealed irrelevant wherever it is not revealed, with $r = s$ and $w = \emptyset$. When every value does, nothing can be latent, and testing deliberation as if no value were latent is without loss.

Example 1 illustrates the second case. Only period 2 reveals gender, and $U_1 \cap U_2 = \{e\}$. At period 2, gender makes no difference on combinations of the degree alone: $\{g\} \sim_2 \emptyset$ and $\{e, g\} \sim_2 \{e\}$. So gender is not revealed irrelevant at period 1; the same holds for the field, and at period 3. The freedom thus lies entirely in the status of g and f at periods 1 and 3. Which values can be latent together is a harder question. Gender and the field can each be latent at period 1, but not both: with both relevant, period 1 would have to rank candidates by u, as period 2 does.

4 Populations

So far, the analysis has concerned a single DM observed over time. Yet the example that opens the paper compares two recruiters. They make the same choices, read the same book, and end up ranking candidates differently. Do they disagree about which values should matter?

The model separates two things that could differ between them. The weighing relation says how a DM weighs the values she deems relevant: in Example 1, the degree counts for more than gender and the field together. The metapreference says which values she thinks should be relevant at all. It is a view about values, of the kind that a book like Sandel's puts into question. Two recruiters who share a metapreference agree on which values should matter, even if they weigh them

on d_k at period k and keep it on: $M'_t = (M_t \setminus D) \cup \{d_k : 3 \leq k \leq t\}$. This path is still rationalizable, since values that affect no choice change no comparison, and it is Deliberate: for $\tau \geq t+2$, d_τ belongs to M'_τ but not to M'_t, and DM does not change it between t and $t+1$ (see Appendix A). Whenever some relevance path is rationalizable, the data are then deliberate. A change in a value that never affects a choice cannot be excluded by choices, so a test keeps its full power only if every value the analyst lists affects some choice. One could instead require such values never to be relevant; they then play no role, and the results hold with $\overline{M}$ in place of $\mathcal{M}$.

differently and start from different relevance sets. Their rankings can still move apart, because what each held latent before reading the book determines what the book can change. Their divergence reflects a disagreement about values only if they share no metapreference. This section asks when the choices of several individuals are consistent with a shared one.

The question also has an empirical advantage. Many panels observe each individual only twice, before and after a campaign or an intervention. With two observations, the Deliberate condition has no content: any change of a single DM that one weighing relation fits can be read as deliberate. A shared metapreference, by contrast, can be rejected with two observations per individual (Section 4.2).

4.1 Common metapreference

Let I be a finite set of individuals, with $|I| \geq 2$. Individual $i \in I$ is observed at $T_i \geq 2$ periods, her preferences $(\succsim_t^i)_{t=1}^{T_i}$ are complete preorders on X, and U_t^i is the set of values revealed relevant for her at period t. From now on, we assume that every value affects some choice of every individual: $\bigcup_t U_t^i = \mathcal{M}$ for all i. This assumption is strong. It excludes, for instance, the second recruiter of our opening example, whose choices never reveal gender; that example only motivates the question. The characterizations below hold for any profile, but the assumption limits how freely a profile can be chosen.[9]

Everything runs as before, individual by individual. Each individual has her own relevance path $(M_t^i)_t$, with $U_t^i \subseteq M_t^i$, her own weighing relation and her own reconsidered values. A *profile* is a family $(M_t^i)_{t,i}$ of relevance paths, one for each individual. It is *rationalizable* if each path is. For $t < T_i$, let

$$R_t^i = R(M_t^i, M_t^i \triangle M_{t+1}^i)$$

[9]For a single DM, values that affect no choice weaken any test of deliberation (footnote 8). In a population, they matter in a further way. A value that affects the choices of others but never those of i could otherwise be added to every relevance set of i without altering any of her changes. This would move her entire path to a region of $2^{\mathcal{M}}$ that others need not visit, and could remove conflicts with them at no cost.

be the relevance sets available to i at t when she reconsiders exactly what she changes.

Definition 6. A profile is *generated by a common metapreference* if a single metapreference $\rhd$ generates every path $(M_t^i)_t$ by deliberation. The population *admits* a common metapreference if some rationalizable profile is generated by one.

A common metapreference constrains only how individuals rank relevance sets. Their weighing relations, their initial relevance sets and the values they reconsider may all differ. Divergence alone therefore does not contradict the existence of a common metapreference: two individuals who reconsider different values can end up far apart. What does contradict it is a comparison made in opposite ways: two individuals whose choices, possibly through a chain of choices over time, rank the same two relevance sets differently. If that is the case, they cannot be maximizing the same metapreference.

Formally, say that individual i has *turned down* a relevance set S by period τ if $S \in R_t^i$ for some $t < \tau$ and $S \neq M_\tau^i$: S was available to her before τ, and at τ she holds another set. If i deliberates, she then ranks M_τ^i above S, whatever metapreference generates her path (Lemma 7 in Appendix A). The contradiction arises if, for some periods τ and σ, an individual i has turned down the set M_σ^j by τ and an individual j has turned down the set M_τ^i by σ. We then say that i and j *cross*.

For two individuals, precluding this pattern is all that a common metapreference adds to the Deliberate condition (Theorem 2 below). With three individuals or more, checking pairs no longer suffices, since there can be Condorcet cycles that no transitive metapreference can accommodate: no two individuals cross, so each pair could share a metapreference, yet the comparisons that their paths reveal run in a circle, and no metapreference serves all three.[10] A common metapreference must therefore

[10]Let $\mathcal{M} = \{a, b, c\}$, $I = \{1, 2, 3\}$ and $T_1 = T_2 = T_3 = 2$. Individual 1 ranks alternatives by how many of b and c they contain at period 1, and by how many of a and c they contain at period 2. Likewise, individual 2 moves from $\{a, c\}$ to $\{a, b\}$, and individual 3 from $\{a, b\}$ to $\{b, c\}$. Preferences are additive and every value is revealed, so by Theorem 1 each individual's only rationalizable path is $(U_t^i)_t$. Individual 1, who keeps c and could reach any set containing it, has turned down $\{b, c\}$, which individual 3 holds at period 2. Likewise, individual 3 has turned down $\{a, b\}$, which individual 2 holds, and individual 2 has turned down $\{a, c\}$, which individual 1 holds. This is a crossing cycle, so by Theorem 2 below the population admits no common metapreference. Yet no two of them cross.

also exclude longer cycles of the same kind.

Definition 7. A *crossing cycle* consists of $n \geq 2$ distinct individuals $i_1, \ldots, i_n$ and periods τ_k such that each i_k has turned down $M^{i_{k+1}}_{\tau_{k+1}}$ by τ_k, with $(i_{n+1}, \tau_{n+1}) = (i_1, \tau_1)$.

Together with the Deliberate condition, the absence of crossing cycles characterizes a common metapreference:

Theorem 2. *A profile is generated by a common metapreference if and only if each path is Deliberate and there is no crossing cycle.*

4.2 Testability with two observations

As noted at the start of this section, the Deliberate condition constrains only periods $\tau \geq t + 2$, so every rationalizable path of a DM observed twice is Deliberate, and her data are deliberate. The existence of a common metapreference, by contrast, can be rejected with two observations per individual. But a crossing is defined on relevance sets, which the analyst does not observe and which latent values leave undetermined. This subsection gives a condition on observed preferences alone, which mentions neither relevance sets nor metapreferences, under which two individuals cross whatever is latent, so that the existence of a common metapreference is rejected (Proposition 2). For two individuals observed twice, the condition is also necessary.

Consider an individual i and two consecutive periods t and $t + 1$ that jointly reveal every value: $U^i_t \cup U^i_{t+1} = \mathcal{M}$. Write $H^i_t = U^i_t \cap U^i_{t+1}$ for the values i reveals at both periods, $L^i_t = U^i_t \setminus U^i_{t+1}$ for those she stops revealing, and $G^i_t = U^i_{t+1} \setminus U^i_t$ for those she starts revealing. These sets partition $\mathcal{M}$. A relevance path says which of these revealed changes are real: a value of L^i_t that belongs to M^i_{t+1} stays latently relevant at $t + 1$, and a value of G^i_t that belongs to M^i_t was already latently relevant at t. Thus

$$M^i_t = H^i_t \cup L^i_t \cup (M^i_t \cap G^i_t), \qquad M^i_{t+1} = H^i_t \cup G^i_t \cup (M^i_{t+1} \cap L^i_t),$$

and the sets of R_t^i are those that contain $M_t^i \cap M_{t+1}^i$. Which revealed changes are real? Let P_t^i be the set of values of L_t^i revealed irrelevant at $t+1$ and of G_t^i revealed irrelevant at t. Their changes are real along every rationalizable path (Lemma 6 in Appendix A): i has truly dropped the former and truly taken up the latter. The other revealed changes may be only apparent. When every value makes a difference on its own at some period, $P_t^i = L_t^i \cup G_t^i$ (Section 3).

Proposition 2. *Let $i \neq j$, and let $t < T_i$ and $s < T_j$ be such that $U_t^i \cup U_{t+1}^i = U_s^j \cup U_{s+1}^j = \mathcal{M}$. Consider the conditions (a) $L_s^j \not\subseteq P_t^i$, (b) $L_t^i \not\subseteq P_s^j$, and (c) $\succsim_t^i = \succsim_{t+1}^i$ on $2^{H_t^i \cup (L_t^i \setminus L_s^j)}$, and $\succsim_s^j = \succsim_{s+1}^j$ on $2^{H_s^j \cup (L_s^j \setminus L_t^i)}$.*

(i) If none of them holds, the population admits no common metapreference.

(ii) If $I = \{i, j\}$, $T_i = T_j = 2$, and each individual has a rationalizable relevance path, the converse holds: if one of the conditions holds, the population admits a common metapreference.

Conditions (a) and (b) say that the final relevance set of one individual may have been out of the other's reach. Under (a), some value that j stops revealing is one that i need not have reconsidered: she reveals it at both periods, or her revealed change in it can be latent. When each is observed twice, condition (c) says that the two may have ended at the same relevance set: each can keep latently relevant what she stops revealing while the other reveals it at the later period. When all three fail, each has turned down the set the other holds, whatever is latent. Part (ii) thus characterizes, for two individuals observed twice whose data are each rationalizable, the data that reject a common metapreference.

The simplest rejection connects the same two relevance sets in opposite directions.

Corollary 2. *Let $i, j \in I$ be distinct, and let $m \in \mathcal{M}$. Suppose that, for some $t < T_i$ and $s < T_j$, $U_t^i = U_{s+1}^j = \mathcal{M} \setminus \{m\}$ and $U_{t+1}^i = U_s^j = \mathcal{M}$. Then the population admits no common metapreference.*

The two individuals move in opposite directions between the same two relevance sets: i takes up m, while j drops it. Neither set can hide a latent value—$\mathcal{M}$

because it contains every value, and $\mathcal{M} \setminus \{m\}$ because m is revealed irrelevant at the periods that do not reveal it—so no latent value can tell the two moves apart. In the recruiting example, consider two recruiters who never attend to anything beyond degrees, skills tests and gender: one, who weighs degrees and skills tests, starts taking gender into account, while the other, who weighs all three, stops doing so. Whatever they attend to without acting on it, they cannot share a metapreference.

Otherwise, latent values can reconcile individuals whose revealed changes are exactly opposite.

Example 2. Let $\mathcal{M} = \{a, b, c\}$, $I = \{1, 2\}$ and $T_1 = T_2 = 2$, and let both individuals' weighing relations be represented by $u(S) = 2 \cdot \mathbf{1}[a \in S] + \mathbf{1}[\{b, c\} \subseteq S]$, the weighing relation of Example 1 with a, b, c in place of e, g, f. At period 1, individual 1 ranks alternatives by whether they contain a, and at period 2 by u; individual 2 does the reverse. Then $U_1^1 = U_2^2 = \{a\}$ and $U_2^1 = U_1^2 = \{a, b, c\}$. If no value is latent, the paths are $(\{a\}, \{a, b, c\})$ and $(\{a, b, c\}, \{a\})$, and the two individuals cross.

Now let individual 1 already regard b as relevant at period 1, and let individual 2 keep c relevant at period 2: the paths are $(\{a, b\}, \{a, b, c\})$ and $(\{a, b, c\}, \{a, c\})$. The weighing relation represented by u rationalizes both, since $u(x \cap \{a, b\}) = u(x \cap \{a, c\}) = 2 \cdot \mathbf{1}[a \in x]$. Individual 2 has turned down the set individual 1 holds at period 2: she held $\{a, b, c\}$ and left it. But individual 1 has not turned down the set individual 2 holds at period 2, since $\{a, c\} \notin R(\{a, b\}, \{c\})$, nor at period 1, since she holds $\{a, b, c\}$ herself at period 2. There is no crossing, and any metapreference with $\{a, c\} \rhd \{a, b, c\} \rhd \{a, b\}$ generates both paths.

In the terms of Proposition 2, neither b nor c is revealed irrelevant at period 1 for individual 1: each alone makes no difference at period 2 on combinations of a, so condition (a) holds. But they cannot both be latent: period 1 would then have to rank $\{a, b, c\}$ above $\{a\}$, as period 2 does. Individual 1 can have regarded one of them as relevant at period 1, but not both. The two moves are opposite only in appearance. Individual 1, who already regards b as relevant, takes up c, and b starts to matter. Individual 2 drops b, and c, which she keeps, stops mattering. One adds c; the other removes b. As in Section 3, the reconciliation rests on an interaction, here

between b and c; Corollary 2 shows that, when the two sets are $\mathcal{M} \setminus \{m\}$ and $\mathcal{M}$, there is no room for one.

Whether one person deliberates depends only on what she changes. Whether two people can share a metapreference also depends on where they start: two individuals who make the same changes from different relevance sets may or may not conflict, and two who connect the same sets in opposite directions always do. Latent values therefore play a second role in populations. For a single DM, they distinguish a later relevance set from an earlier one. In a population, they also determine which sets a change connects, and hence whether the changes of two individuals conflict at all.

With three individuals or more, the conditions of Proposition 2 remain necessary for each pair, but they are no longer sufficient. Each individual's relevance path must serve every pair at once, and a common metapreference must also exclude crossing cycles (footnote 10).

5 Conclusion

Choices reveal which values a decision maker uses, not all the values she holds. We have shown when this gap closes, so that it cannot affect the identification of deliberate change. A value can be latent at a period only if the data do not reveal it irrelevant there, which, for a value that affects some choice, requires its effect to depend on values the decision maker then ignores (Section 3). When, at every period, every value is revealed relevant or revealed irrelevant, testing deliberation as if nothing were latent is without loss. Otherwise, as for the first recruiter, latent values can turn an apparent reversal into a deliberate change. In a population, the question becomes whether individuals share a view of which values should matter. They can exactly when some rationalizable paths are Deliberate and form no crossing cycle (Theorem 2), and two observations per individual can be enough to reject such a view (Proposition 2).

Latent values also call for a probabilistic model. When the data leave several

relevance paths open, the analyst cannot say whether a change is deliberate, but she can say how likely it is. She can assign probabilities to the values that may be latent at each period, over the rationalizable paths only, and revise them as the decision maker's preferences change: each new period may reveal some values irrelevant and rule out the paths that made them relevant. By Corollary 1, the probability that the data are deliberate is then the probability of the rationalizable paths that are Deliberate. Under the condition of Theorem 1, this probability is zero or one whatever the analyst's prior. The converse does not hold, since several rationalizable paths may all be Deliberate, or none. Elsewhere, it measures how much the verdict owes to what she assumes to be latent. In a population, the same updating would bear on the profiles compatible with a common metapreference.

Statements and Declarations

Funding. The author acknowledges support from the project JICE-InChangE, funded by the programme "Profilbildung 2020" of the Ministry of Culture and Science of the State of North Rhine-Westphalia.

Competing interests. The author has no competing interests to declare that are relevant to the content of this article.

Data availability. Not applicable. This article is purely theoretical, and no datasets were generated or analysed.

Use of generative AI. This paper builds upon a more general version completed in 2022/2023. The author used Claude (Anthropic) to assist with verification and exposition. Specifically, AI was used to: (i) verify mathematical detail in theorems and proofs, checking logical structure and catching imprecision in stated hypotheses; (ii) check the numerical examples; (iii) improve exposition, notation consistency, and formatting; (iv) migrate the bibliography to BibTeX; and (v) develop, at the author's direction, the formal statements and proofs of the population analysis of Section 4,

including Lemma 9 and Proposition 2, and of the identification results of Section 3. The model, the research questions and the direction of the analysis are the author's. The author retains full responsibility for correctness and interpretation.

A Proofs

Throughout, for $M, S, A \subseteq \mathcal{M}$, $S \in R(M, A)$ if and only if $M \triangle S \subseteq A$. In particular, $M \in R(M, A)$ for every A, and $R(M, A) \subseteq R(M, A')$ whenever $A \subseteq A'$.

A.1 Proofs for Section 2

Lemma 2. *Let t be a period and $M \supseteq U_t$. (a) If $x \cap M = y \cap M$, then $x \sim_t y$. (b) A pair $(M, \geqslant)$ rationalizes $\succsim_t$ if and only if $\geqslant$ coincides with $\succsim_t$ on 2^M.*

Proof. (a) Write $x \triangle y = \{c_1, \ldots, c_p\}$ and $z_k = x \triangle \{c_1, \ldots, c_k\}$. Then $z_0 = x$, $z_p = y$, and z_{k-1} and z_k differ only in c_k, which lies outside $M \supseteq U_t$ since $x \cap M = y \cap M$. Hence $z_{k-1} \sim_t z_k$ for every k, and $x \sim_t y$ by transitivity of $\sim_t$.

(b) If $(M, \geqslant)$ rationalizes $\succsim_t$ and $S, S' \subseteq M$, then $S \succsim_t S' \iff S \cap M \geqslant S' \cap M \iff S \geqslant S'$. Conversely, if $\geqslant$ coincides with $\succsim_t$ on 2^M, then by (a), since x and $x \cap M$ coincide on M, $x \succsim_t y \iff x \cap M \succsim_t y \cap M \iff x \cap M \geqslant y \cap M$. □

Proof of Lemma 1. If $(M, \geqslant)$ rationalizes $\succsim_t$ and x, y differ only in some $m \notin M$, then $x \cap M = y \cap M$, so $x \sim_t y$, since the complete relation $\geqslant$ is reflexive; hence $U_t \subseteq M$. Conversely, if $U_t \subseteq M$, then $(M, \succsim_t)$ rationalizes $\succsim_t$ by Lemma 2(b). □

Lemma 3. *A relevance path $(M_t)_t$ is rationalizable if and only if, for all periods s and t, $\succsim_s$ and $\succsim_t$ coincide on $2^{M_s \cap M_t}$.*

Proof. Necessity. If a weighing relation $\geqslant$ rationalizes every period along the path, Lemma 2(b) shows that $\geqslant$ coincides with $\succsim_s$ on 2^{M_s} and with $\succsim_t$ on 2^{M_t}, so $\succsim_s$ and $\succsim_t$ coincide on $2^{M_s \cap M_t}$.

Sufficiency. Let $S \geqslant S'$ if $S \succsim_r S'$ for every r with $S, S' \subseteq M_r$. By the condition, all such r agree, so $\geqslant$ is complete and coincides with $\succsim_t$ on 2^{M_t} for every t; Lemma 2(b) concludes. □

Lemma 4. *Let $\rhd$ be a strict partial order on $2^{\mathcal{M}}$, with reflexive closure $\trianglerighteq$. Suppose that, for some $(A_t)_{t<T}$ and all $t < T$, $M_{t+1} \in R(M_t, A_t)$ and $M_{t+1} \rhd S$ for every other $S \in R(M_t, A_t)$. Then $M_\tau \trianglerighteq M_{t+1} \trianglerighteq S$ whenever $t < \tau$ and $S \in R(M_t, M_t \triangle M_{t+1})$.*

Proof. The relation $\trianglerighteq$ is a partial order. For every $u < T$, $M_u \in R(M_u, A_u)$, so $M_{u+1} \trianglerighteq M_u$; chaining gives $M_\tau \trianglerighteq M_{t+1}$ whenever $t < \tau$. Moreover, $M_{t+1} \in R(M_t, A_t)$ gives $M_t \triangle M_{t+1} \subseteq A_t$, so $R(M_t, M_t \triangle M_{t+1}) \subseteq R(M_t, A_t)$, and $M_{t+1} \trianglerighteq S$ for every S in it. □

Proof of Proposition 1. By the identity above, the premise of the Deliberate condition reads $M_\tau \in R(M_t, M_t \triangle M_{t+1})$.

Only if. Let $\rhd$ and $(A_t)_t$ generate the sequence, and let $\tau \geq t + 2$ with $M_\tau \in R(M_t, M_t \triangle M_{t+1})$. Lemma 4 gives $M_\tau \trianglerighteq M_{t+1} \trianglerighteq M_\tau$, hence $M_\tau = M_{t+1}$.

If. Suppose that the sequence is Deliberate. We construct a metapreference that generates it with $A_t = M_t \triangle M_{t+1}$ for all $t < T$. For a set M that appears in the sequence, let $\ell(M) = \max\{t : M_t = M\}$ be the last period at which it appears; distinct sets have distinct last periods. Let $\rhd$ rank the sets that appear in the sequence by ℓ, later being better, above all other subsets of $\mathcal{M}$, which are ordered arbitrarily. Fix $t < T$ and $M \in R(M_t, M_t \triangle M_{t+1}) \setminus \{M_{t+1}\}$. If M does not appear in the sequence, $M_{t+1} \rhd M$. Otherwise, suppose by contradiction that $\tau := \ell(M) > \ell(M_{t+1}) \geq t + 1$. Then $\tau \geq t + 2$ and $M_\tau = M \in R(M_t, M_t \triangle M_{t+1})$, so the Deliberate condition gives $M = M_{t+1}$, a contradiction. Hence $\ell(M) < \ell(M_{t+1})$ and $M_{t+1} \rhd M$. Since $M_{t+1} \in R(M_t, M_t \triangle M_{t+1})$, $M_{t+1} = \max(R(M_t, M_t \triangle M_{t+1}), \rhd)$, and the sequence is generated by deliberation. □

Proof of Corollary 1. If the data are deliberate, with relevance sets $(M_t)_t$ and weighing relation $\geqslant$, Lemma 1 gives $U_t \subseteq M_t$ for every t, so $(M_t)_t$ is a rationalizable relevance path, and it is Deliberate by Proposition 1. Conversely, if a rationalizable relevance path is Deliberate, Proposition 1 shows that it is generated by deliberation, and the data are deliberate. □

A.2 Proofs for Section 3

Lemma 5. *Let $(M_t)_t$ be a relevance path rationalized by $\succcurlyeq$. Then $\succcurlyeq$ also rationalizes every relevance path $(M'_t)_t$ with $M'_t \subseteq M_t$ for all t.*

Proof. Since $\succcurlyeq$ rationalizes $(M_t)_t$, it coincides with $\succsim_t$ on $2^{M_t} \supseteq 2^{M'_t}$ for every t, so it rationalizes $(M'_t)_t$ by Lemma 2(b). □

Lemma 6. *Let $(M_r)_r$ be a rationalizable relevance path. If m is revealed irrelevant at t, then $m \notin M_t$.*

Proof. Since m is revealed irrelevant at t, there are r and w with $m \in U_r$, $w \subseteq U_t \cap U_r$ and $w \cup \{m\} \not\sim_r w$. Suppose that $m \in M_t$. Then $w \cup \{m\}$ and w are subsets of $M_t \cap M_r$, since $U_t \subseteq M_t$ and $U_r \subseteq M_r$. Yet $w \cup \{m\} \sim_t w$, since the two differ only in $m \notin U_t$. This contradicts Lemma 3. □

Proof of Theorem 1. For a period t and a value $m \notin U_t$, let $(M_r^{t,m})_r$ be the relevance path that adds m to U_t only: $M_t^{t,m} = U_t \cup \{m\}$ and $M_r^{t,m} = U_r$ for $r \neq t$. The intersections of its distinct periods are those of $(U_r)_r$, except that $M_t^{t,m} \cap M_r^{t,m} = (U_t \cap U_r) \cup \{m\}$ when $m \in U_r$.

Only if. Suppose that $(U_t)_t$ is the only rationalizable relevance path, and let t be a period and $m \in \mathcal{M} \setminus U_t$. The path $(M_r^{t,m})_r$ is not rationalizable, so by Lemma 3 two of its periods disagree on the combinations of values relevant at both. This cannot happen for a pair not involving t, nor for t paired with any r such that $m \notin U_r$, since $(U_r)_r$ is rationalizable. So there is r with $m \in U_r$ such that $\succsim_t$ and $\succsim_r$ do not coincide on $2^{(U_t \cap U_r) \cup \{m\}}$. Suppose, by contradiction, that $w \cup \{m\} \sim_r w$ for every $w \subseteq U_t \cap U_r$. Then every $x \subseteq (U_t \cap U_r) \cup \{m\}$ satisfies $x \sim_r x \setminus \{m\}$, and $x \sim_t x \setminus \{m\}$ since the two differ only in $m \notin U_t$. As $\succsim_t$ and $\succsim_r$ coincide on $2^{U_t \cap U_r}$, for all $x, y \subseteq (U_t \cap U_r) \cup \{m\}$,

$$x \succsim_t y \iff x \setminus \{m\} \succsim_t y \setminus \{m\} \iff x \setminus \{m\} \succsim_r y \setminus \{m\} \iff x \succsim_r y,$$

a contradiction. Hence m is revealed irrelevant at t.

If. Let $(M_t)_t$ be a rationalizable relevance path, and suppose that some $m \in M_t$ is not in U_t. By assumption, m is revealed irrelevant at t, and Lemma 6 gives $m \notin M_t$, a contradiction. Hence $M_t = U_t$ for every t. □

Proof of the claim in footnote 8. Let $\geqslant$ rationalize $(M_t)_t$. Since no value of D is revealed, $(M_t \setminus D)_t$ is a relevance path, and $\geqslant$ rationalizes it by Lemma 5. Let $S \geqslant' S'$ if $S \setminus D \geqslant S' \setminus D$. Since $(x \cap M'_t) \setminus D = x \cap (M_t \setminus D)$, the pair $(M'_t, \geqslant')$ rationalizes $\succsim_t$ for every t. For the Deliberate condition, let $\tau \geq t+2$. Then $d_\tau \in M'_\tau \setminus M'_t$, while $d_\tau \notin M'_t \triangle M'_{t+1}$, since d_τ belongs to neither M'_t nor M'_{t+1}. Hence $M'_t \triangle M'_\tau \not\subseteq M'_t \triangle M'_{t+1}$, and the path is Deliberate. By Corollary 1, the data are deliberate. □

A.3 Proofs for Section 4

Throughout, $(M^i_t)_{t,i}$ is a profile and R^i_t is as in Section 4. Let C^i be the relation on $2^{\mathcal{M}}$ defined by $S\ C^i\ S'$ if $S = M^i_\tau$ for some τ, $S' \in R^i_t$ for some $t < \tau$, and $S' \neq S$. If $S\ C^i\ S'$, the period τ can be taken to be the last period at which i holds S. If i has turned down S by τ, then $M^i_\tau\ C^i\ S$; conversely, if $M^i_\tau\ C^i\ S$ and τ is the last period at which i holds M^i_τ, then i has turned down S by τ.

Lemma 7. *Suppose that a metapreference $\rhd$ generates the path of i by deliberation. If i has turned down S by τ, then $M^i_\tau \rhd S$.*

Proof. Let $\rhd$ and $(A^i_t)_t$ generate the path of i, and let $S \in R^i_t$ with $t < \tau$ and $S \neq M^i_\tau$. Lemma 4 gives $M^i_\tau \trianglerighteq M^i_{t+1} \trianglerighteq S$, hence $M^i_\tau \rhd S$. □

Lemma 8. *If the path of i is Deliberate, then C^i is a strict partial order, and $M^i_{t+1}\ C^i\ S$ for all $t < T_i$ and $S \in R^i_t \setminus \{M^i_{t+1}\}$.*

Proof. Drop the superscript i. The last statement holds by definition, with $\tau = t+1$, and C is irreflexive by definition. The Deliberate condition has two consequences. First, for $t < \tau$, $M_\tau \in R_t$ implies $M_\tau = M_{t+1}$: this is trivial for $\tau = t+1$, and it is the Deliberate condition for $\tau \geq t+2$. Second, if $M_a = M_c$ and $a \leq b \leq c$, then $M_b = M_a$. Indeed, suppose $M_b \neq M_a$, and let u be the last period in $\{a, \ldots, b-1\}$ with $M_u = M_a$. Then $M_{u+1} \neq M_a$, so $c \geq u+2$. Since $M_c = M_u \in R_u$, the Deliberate condition gives $M_c = M_{u+1}$, a contradiction.

Now let $S_1\ C\ S_2\ C\ S_3$, with $S_1 = M_\tau$, $S_2 \in R_t$ for some $t < \tau$, $S_2 = M_{\tau'}$, $S_3 \in R_{t'}$ for some $t' < \tau'$, $S_2 \neq S_1$ and $S_3 \neq S_2$. We claim that $\tau' < \tau$. Otherwise $\tau' \geq \tau > t$,

so $M_{\tau'} \in R_t$ gives $M_{\tau'} = M_{t+1}$, and since $t+1 \leq \tau \leq \tau'$, $M_\tau = M_{\tau'}$, that is, $S_1 = S_2$, a contradiction. Hence $t' < \tau$ and $S_3 \in R_{t'}$. Suppose $S_3 = S_1 = M_\tau$. Then $M_\tau \in R_{t'}$ with $t' < \tau$, so $M_\tau = M_{t'+1}$, and since $t'+1 \leq \tau' < \tau$, $M_{\tau'} = M_\tau$, that is, $S_2 = S_1$, a contradiction. Hence $S_1\ C\ S_3$, with the same period τ for S_1, and C is transitive. □

Proof of Theorem 2. Only if. Let $\rhd$ generate every path. Each path is Deliberate by Proposition 1. Suppose that there is a crossing cycle. For each k, i_k has turned down $M^{i_{k+1}}_{\tau_{k+1}}$ by τ_k, so Lemma 7 gives $M^{i_k}_{\tau_k} \rhd M^{i_{k+1}}_{\tau_{k+1}}$. Chaining these rankings around the cycle gives $M^{i_1}_{\tau_1} \rhd M^{i_1}_{\tau_1}$, a contradiction.

If. Suppose that each path is Deliberate and that there is no crossing cycle. We construct a common metapreference that generates each path with $A^i_t = M^i_t \triangle M^i_{t+1}$. Let $Q = \bigcup_i C^i$. By Lemma 8, each C^i is a strict partial order.

Step 1: Q is acyclic. Suppose not, and take a cycle of Q of minimal length n: sets $S_1, \ldots, S_n$ and individuals $i_1, \ldots, i_n$ with $S_k\ C^{i_k}\ S_{k+1}$ for all k, indices being taken modulo n. Call $S_k\ C^{i_k}\ S_{k+1}$ link k, and say that i_k owns it. Since each C^i is irreflexive, $n \geq 2$. Let τ_k be the last period at which i_k holds S_k, so that $M^{i_k}_{\tau_k} = S_k$.

We first show that no individual owns two links. Suppose that i does. Among the links she owns, let a be one with τ_a largest, and let $b \neq a$ be another. Then $S_a\ C^i\ S_{b+1}$. If $S_b = S_a$, this is link b. Otherwise $\tau_b \neq \tau_a$, since i holds different sets at these periods, so $\tau_b < \tau_a$. As $S_b = M^i_{\tau_b} \in R^i_{\tau_b}$, the definition of C^i, with $\tau = \tau_a$ and $t = \tau_b$, gives $S_a\ C^i\ S_b$. Link b and the transitivity of C^i then give $S_a\ C^i\ S_{b+1}$.

Now follow the cycle from S_a: the links $a, a+1, \ldots, b$ lead to S_{b+1}, and there are at least two of them since $b \neq a$. By irreflexivity, $S_{b+1} \neq S_a$, so these links are not the whole cycle. Replacing them by the single link $S_a\ C^i\ S_{b+1}$ therefore gives a strictly shorter cycle of Q, a contradiction.

Hence the individuals i_k are distinct. For each k, link k says that $S_{k+1} \in R^{i_k}_t$ for some $t < \tau_k$ and that $S_{k+1} \neq S_k = M^{i_k}_{\tau_k}$. So i_k has turned down $S_{k+1} = M^{i_{k+1}}_{\tau_{k+1}}$ by τ_k. This is a crossing cycle, which is excluded.

Step 2: the metapreference. Since Q is acyclic, its transitive closure is a strict partial order, which extends to a linear order $\rhd$ (Szpilrajn, 1930). Fix i and $t < T_i$, and let $A^i_t = M^i_t \triangle M^i_{t+1}$, so that $R(M^i_t, A^i_t) = R^i_t$. Then $M^i_{t+1} \in R^i_t$, and by Lemma 8,

$M^i_{t+1}\ C^i\ S$, hence $M^i_{t+1} \rhd S$, for every other $S \in R^i_t$. Thus $M^i_{t+1} = \max(R^i_t, \rhd)$, and $\rhd$ generates the path of i. □

For an individual i and a period t with $U^i_t \cup U^i_{t+1} = \mathcal{M}$, say that a set $B \subseteq L^i_t \cup G^i_t$ *can stay latent* for (i,t) if $x \sim^i_t x \cap H^i_t$ for every $x \subseteq H^i_t \cup (B \cap L^i_t)$, and $x \sim^i_{t+1} x \cap H^i_t$ for every $x \subseteq H^i_t \cup (B \cap G^i_t)$. Such sets are closed under subsets.

Lemma 9. *Let $U^i_t \cup U^i_{t+1} = \mathcal{M}$, and let $M^i_t \supseteq U^i_t$ and $M^i_{t+1} \supseteq U^i_{t+1}$. Then $\succsim^i_t$ and $\succsim^i_{t+1}$ coincide on $2^{M^i_t \cap M^i_{t+1}}$ if and only if they coincide on $2^{H^i_t}$ and $(M^i_{t+1} \cap L^i_t) \cup (M^i_t \cap G^i_t)$ can stay latent.*

Proof. Drop the superscript i and the subscript t of H, L and G. Since $H \subseteq M_t \cap M_{t+1}$, $L \subseteq M_t$ and $G \subseteq M_{t+1}$,

$$W := M_t \cap M_{t+1} = H \cup K \cup N, \qquad K = M_{t+1} \cap L, \quad N = M_t \cap G.$$

For $x \subseteq W$, $x \cap U_t = x \cap (H \cup K)$ and $x \cap U_{t+1} = x \cap (H \cup N)$, so Lemma 2(a) gives $x \sim_t x \cap (H \cup K)$ and $x \sim_{t+1} x \cap (H \cup N)$.

If. Let $x, y \subseteq W$. Since $K \cup N$ can stay latent, $x \cap (H \cup K) \sim_t x \cap H$, and likewise for y, and $x \cap (H \cup N) \sim_{t+1} x \cap H$, and likewise for y. Hence

$$x \succsim_t y \iff x \cap H \succsim_t y \cap H \iff x \cap H \succsim_{t+1} y \cap H \iff x \succsim_{t+1} y,$$

where the middle equivalence uses that the two preferences coincide on 2^H.

Only if. The two preferences coincide on 2^H because $2^H \subseteq 2^W$. Let $x \subseteq H \cup K$. Then x and $x \cap H$ have the same trace on $U_{t+1} = H \cup G$, so $x \sim_{t+1} x \cap H$ by Lemma 2(a). Both lie in 2^W, so $x \sim_t x \cap H$. The argument for N is symmetric, so $K \cup N$ can stay latent. □

Proof of Proposition 2. Consider a rationalizable relevance path of i. Since $M^i_t \cup M^i_{t+1} \supseteq U^i_t \cup U^i_{t+1} = \mathcal{M}$, a set agrees with M^i_t outside $M^i_t \triangle M^i_{t+1}$ if and only if it contains $M^i_t \cap M^i_{t+1}$, so R^i_t consists of the sets that contain $M^i_t \cap M^i_{t+1}$. By Lemma 6, no value of P^i_t belongs to $M^i_t \cap M^i_{t+1}$. Finally, since $U^i_{t+1} = \mathcal{M} \setminus L^i_t$, $M^i_{t+1} = \mathcal{M} \setminus (L^i_t \setminus M^i_{t+1})$.

(i) Suppose that none of (a)–(c) holds, and take any rationalizable profile. Since (a) fails, $L^j_s \subseteq P^i_t$, so $M^i_t \cap M^i_{t+1}$ is disjoint from L^j_s and is contained in $\mathcal{M} \setminus L^j_s = U^j_{s+1} \subseteq M^j_{s+1}$: hence $M^j_{s+1} \in R^i_t$. Since (b) fails, $M^i_{t+1} \in R^j_s$ likewise. Suppose that $M^i_{t+1} = M^j_{s+1}$. Then $L^i_t \setminus M^i_{t+1} = L^j_s \setminus M^j_{s+1}$, and this set is contained in $L^i_t \cap L^j_s$. Hence $L^i_t \setminus L^j_s \subseteq L^i_t \cap M^i_{t+1}$, and $H^i_t \cup (L^i_t \setminus L^j_s) \subseteq M^i_t \cap M^i_{t+1}$. By Lemma 3, $\succsim^i_t$ and $\succsim^i_{t+1}$ coincide on $2^{M^i_t \cap M^i_{t+1}}$, hence on $2^{H^i_t \cup (L^i_t \setminus L^j_s)}$. Symmetrically for j, so (c) holds, a contradiction. Hence $M^i_{t+1} \neq M^j_{s+1}$: i has turned down M^j_{s+1} by $t+1$, and j has turned down M^i_{t+1} by $s+1$. The two cross, and Theorem 2 concludes.

(ii) Let $I = \{i, j\}$ and $T_i = T_j = 2$, so that $t = s = 1$, and drop these subscripts. Since i has a rationalizable relevance path, (U^i_1, U^i_2) is rationalizable by Lemma 5, so $\succsim^i_1$ and $\succsim^i_2$ coincide on 2^{H^i} by Lemma 3. Hence, by Lemma 9 and Lemma 3, the path $(U^i_1 \cup N, U^i_2 \cup K)$ is rationalizable whenever $K \subseteq L^i$, $N \subseteq G^i$ and $K \cup N$ can stay latent; and likewise for j. With two periods, every path is Deliberate, a crossing cycle is a crossing, and individuals can only have turned down each other's sets by period 2. By Theorem 2, it suffices to find such a profile under which $M^j_2 \notin R^i_1$, or $M^i_2 \notin R^j_1$, or $M^i_2 = M^j_2$.

Under (a), let $m \in L^j \setminus P^i$, and let the path of j be (U^j_1, U^j_2). With two periods, a value m of $L^i \cup G^i$ belongs to P^i if and only if $\{m\}$ cannot stay latent: both say that $x \not\sim x \setminus \{m\}$ for some $x \subseteq H^i \cup \{m\}$ at the period that reveals m. So if $m \in L^i$, the path $(U^i_1, U^i_2 \cup \{m\})$ is rationalizable; if $m \in G^i$, so is $(U^i_1 \cup \{m\}, U^i_2)$; and if $m \in H^i$, let the path of i be (U^i_1, U^i_2). In each case $m \in M^i_1 \cap M^i_2$, so every set of R^i_1 contains m, while $m \notin M^j_2 = U^j_2$: $M^j_2 \notin R^i_1$. Case (b) is symmetric. Under (c), let the paths be $(U^i_1, U^i_2 \cup (L^i \setminus L^j))$ and $(U^j_1, U^j_2 \cup (L^j \setminus L^i))$, which are rationalizable by (c) and Lemma 3. Then $M^i_2 = \mathcal{M} \setminus (L^i \cap L^j) = M^j_2$. □

Proof of Corollary 2. Here $L^i_t = \emptyset$, $G^i_t = L^j_s = \{m\}$ and $H^j_s = \mathcal{M} \setminus \{m\}$. Since m is revealed at $t+1$, some $w \subseteq \mathcal{M} \setminus \{m\} = U^i_t \cap U^i_{t+1}$ satisfies $w \cup \{m\} \not\sim^i_{t+1} w$, so $m \in P^i_t$ and (a) fails. Condition (b) fails since $L^i_t = \emptyset$. Since $m \in U^j_s \setminus U^j_{s+1}$, some $x \ni m$ satisfies $x \not\sim^j_s x \setminus \{m\}$ while $x \sim^j_{s+1} x \setminus \{m\}$, so $\succsim^j_s$ and $\succsim^j_{s+1}$ do not coincide on $2^{H^j_s \cup \{m\}} = 2^{\mathcal{M}}$, and (c) fails. Proposition 2(i) concludes. □

B Transitive weighing relations

Lemma 3 does not require the weighing relation to be transitive. Since $\geqslant$ coincides with $\succsim_t$ on 2^{M_t}, it is transitive on the combinations DM compares at any single period; transitivity on all of Ω only restricts how comparisons made at different periods fit together. Boissonnet et al. (2023) do not require it either, and they note that, with binary attributes, transitive preferences at each period do not ensure it; the transitive case appears in an earlier version of their paper (Boissonnet et al., 2022). This appendix shows that whether a transitive weighing relation fits the data depends only on how DM ranks combinations of the values she reveals at each period, not on which values are latent, and gives a condition under which one always does.

Pairs of periods no longer suffice. Let $\mathcal{M} = \{a_1, a_2, a_3\}$ and suppose that, at period k, DM ranks alternatives by $2 \cdot \mathbf{1}[a_k \in x] + \mathbf{1}[a_{k+1} \in x]$, with $a_4 = a_1$. Any two periods share a single revealed value and rank it in the same way, so $(U_t)_t$ is rationalizable. Yet $\{a_1\} \succ_1 \{a_2\}$, $\{a_2\} \succ_2 \{a_3\}$ and $\{a_3\} \succ_3 \{a_1\}$, and since each value is revealed at the two periods that compare it, these comparisons bind along every relevance path. No transitive weighing relation rationalizes the three periods. The contradiction is spread over three periods, no two of which disagree.

Definition 8. A relevance path $(M_t)_t$ is *acyclic* if there are no combinations $S_1, \ldots, S_n$, with $S_{n+1} = S_1$, and periods $r_1, \ldots, r_n$ such that, for all $k \leq n$, $S_k, S_{k+1} \subseteq M_{r_k}$ and $S_k \succsim_{r_k} S_{k+1}$, with $S_k \succ_{r_k} S_{k+1}$ for some k.

Acyclicity says that a chain of comparisons, each made at some period, cannot return to its start with a strict link. With $n = 2$, it is the condition of Lemma 3.

Proposition 3. *A relevance path is rationalizable by a transitive weighing relation if and only if it is acyclic.*

Lemma 10. *A rationalizable relevance path is acyclic if and only if $(U_t)_t$ is acyclic.*

The analyst can therefore check transitivity on the path $(U_t)_t$ alone: the data are deliberate with a transitive weighing relation if and only if they are deliberate and $(U_t)_t$ is acyclic. If no value is latent, this is the characterization of Boissonnet et al.

(2022, Proposition 3); when $(U_t)_t$ is acyclic, the results of Section 3 hold verbatim for transitive weighing relations.

When does acyclicity come for free? The example above owes its cycle to the way the revealed sets overlap: each pair of periods shares a value, and no period contains what the other two share. The following condition rules this out. It requires the overlaps to be tree-like, as in acyclic hypergraphs (Beeri et al., 1983).

Definition 9. The revealed sets satisfy *Running Intersection* if the periods can be ordered as $t_1, \ldots, t_T$ so that, for every $k \geq 2$, there is $j < k$ with

$$U_{t_k} \cap \left(U_{t_1} \cup \cdots \cup U_{t_{k-1}}\right) \subseteq U_{t_j}.$$

The order need not be the temporal one. The condition holds whenever $T = 2$, when the revealed sets are nested, when some period reveals everything that any two periods share, and when no value returns ($U_t \cap U_{t''} \subseteq U_{t'}$ for all $t < t' < t''$).

Proposition 4. *If the revealed sets satisfy Running Intersection, every rationalizable relevance path is rationalizable by a transitive weighing relation.*

By Lemma 10, it suffices to check that $(U_t)_t$ is acyclic. The condition is sufficient but not necessary: the revealed sets of the example above violate it, yet if DM ranks alternatives at period k by the restriction of $3 \cdot \mathbf{1}[a_1 \in x] + 2 \cdot \mathbf{1}[a_2 \in x] + \mathbf{1}[a_3 \in x]$ to $\{a_k, a_{k+1}\}$, the revealed sets are the same and a single transitive weighing relation rationalizes all three periods. Whether the condition is necessary for every sequence of preferences with given revealed sets, as the analogous result for relational databases suggests (Beeri et al., 1983), we leave open.

B.1 Proofs

For a relevance path $(M_t)_t$, let $\overline{\succsim}_t$ denote the restriction of $\succsim_t$ to 2^{M_t}, with asymmetric part $\overline{\succ}_t$. A *cycle along* $(M_t)_t$ is a sequence as in the definition of acyclicity: combinations $S_1, \ldots, S_n$, with $S_{n+1} = S_1$, and periods $r_1, \ldots, r_n$ such that $S_k, S_{k+1} \subseteq M_{r_k}$ and $S_k \succsim_{r_k} S_{k+1}$ for all k, with a strict preference for some k.

Proof of Proposition 3. By Lemma 2(b), $(M_t, \geqslant)$ rationalizes $\succsim_t$ if and only if $\geqslant$ coincides with $\overline{\succ}_t$ on 2^{M_t}. If a transitive $\geqslant$ rationalizes every period and there is a cycle along the path, then $S_k \geqslant S_{k+1}$ for all k, with $S_k > S_{k+1}$ for some k, while transitivity along the other links gives $S_{k+1} \geqslant S_k$, a contradiction.

Conversely, suppose that the path is acyclic, and let $Q = \bigcup_t \overline{\succ}_t$, a relation on Ω with asymmetric part $P(Q)$. Acyclicity with $n = 2$ gives $\overline{\succ}_t \subseteq P(Q)$ for every t, and the converse inclusion holds by definition, so $P(Q) = \bigcup_t \overline{\succ}_t$. Acyclicity then says that no cycle of Q contains a link in $P(Q)$: Q is Suzumura consistent. By Suzumura (1976), Q has an ordering extension, a complete preorder $\geqslant$ on Ω with $Q \subseteq \geqslant$ and $P(Q) \subseteq >$. Let $S, S' \subseteq M_t$. If $S \overline{\succsim}_t S'$, then $S \geqslant S'$. Otherwise $S' \overline{\succ}_t S$ by completeness of $\overline{\succsim}_t$, so $S' > S$ and not $S \geqslant S'$. Hence $\geqslant$ coincides with $\overline{\succsim}_t$ on 2^{M_t} for every t. □

Proof of Lemma 10. Let $(M_t)_t$ be a rationalizable relevance path. Since $U_t \subseteq M_t$ for every t, a cycle along $(U_t)_t$ is a cycle along $(M_t)_t$, so if $(M_t)_t$ is acyclic, so is $(U_t)_t$. For the converse, recall from Lemma 3 that $\succsim_s$ and $\succsim_t$ coincide on $2^{M_s \cap M_t}$ for all s, t.

Step 1: links can be reduced to values revealed at both periods. Let s, t be periods and $S \subseteq M_s \cap M_t$. We claim that $S \sim_s S \cap U_s \cap U_t$ and $S \sim_t S \cap U_s \cap U_t$. By Lemma 2(a), $S \sim_s S \cap U_s$ and $S \cap U_s \sim_t S \cap U_s \cap U_t$. The last two sets lie in $2^{M_s \cap M_t}$, where $\succsim_s$ and $\succsim_t$ coincide, so $S \cap U_s \sim_s S \cap U_s \cap U_t$, and $S \sim_s S \cap U_s \cap U_t$. Exchanging s and t gives the second claim.

Step 2: a cycle along $(M_t)_t$ yields a cycle along $(U_t)_t$. Let $S_1, \ldots, S_n$ and $r_1, \ldots, r_n$ form a cycle along $(M_t)_t$, with $r_0 = r_n$. Then $S_k \subseteq M_{r_{k-1}} \cap M_{r_k}$ for all k. Let $S'_k = S_k \cap U_{r_{k-1}} \cap U_{r_k}$, with $S'_{n+1} = S'_1$. By Step 1, $S_k \sim_{r_k} S'_k$ and $S_{k+1} \sim_{r_k} S'_{k+1}$, so $S'_k \succsim_{r_k} S'_{k+1}$ for all k, with strictness preserved. Since S'_k and S'_{k+1} are subsets of U_{r_k}, this is a cycle along $(U_t)_t$. □

Proof of Proposition 4. Let $(M_t)_t$ be a rationalizable relevance path. By Lemma 5, $(U_t)_t$ is rationalizable, so $\succsim_s$ and $\succsim_t$ coincide on $2^{U_s \cap U_t}$ for all s, t. By Lemma 10 and Proposition 3, it suffices to show that there is no cycle along $(U_t)_t$.

Relabel the periods as $1, \ldots, T$ in the order given by Running Intersection. We show by induction on k that no cycle along $(U_t)_t$ uses only periods in $\{1, \ldots, k\}$. For $k = 1$, every link is a comparison under $\succsim_1$, and transitivity excludes a strict link.

Let $k \geq 2$ and consider a cycle along $(U_t)_t$ that uses only periods in $\{1, \ldots, k\}$. If every link is at period k, transitivity of $\succsim_k$ excludes a strict link. Otherwise, take a maximal run of consecutive links at period k, say from S_p to S_q, so that the link into S_p is at some period $r < k$ and the link out of S_q at some period $r' < k$. By transitivity of $\succsim_k$, $S_p \succsim_k S_q$, strictly if some link of the run is strict. Moreover $S_p \subseteq U_r \cap U_k$ and $S_q \subseteq U_k \cap U_{r'}$, so both lie in $U_k \cap (U_1 \cup \cdots \cup U_{k-1}) \subseteq U_{j^*}$ for the period $j^* < k$ given by Running Intersection. As $S_p, S_q \subseteq U_k \cap U_{j^*}$, $S_p \succsim_{j^*} S_q$, with the same strictness. Replace the run by this single link at j^*. The result is still a cycle along $(U_t)_t$, since $S_p, S_q \subseteq U_{j^*}$, and it has a strict link. Doing so for every run of links at period k yields a cycle along $(U_t)_t$ that uses only periods in $\{1, \ldots, k-1\}$, which the induction hypothesis excludes. □

References

Catriel Beeri, Ronald Fagin, David Maier, and Mihalis Yannakakis. On the desirability of acyclic database schemes. *Journal of the ACM*, 30(3):479–513, July 1983. doi: 10.1145/2402.322389.

B. Douglas Bernheim, Luca Braghieri, Alejandro Martínez-Marquina, and David Zuckerman. A theory of chosen preferences. *American Economic Review*, 111(2): 720–754, 2021. doi: 10.1257/aer.20190390.

Niels Boissonnet, Alexis Ghersengorin, and Simon Gleyze. Revealed deliberate preference change. Working paper, HAL hal-03672734, 2022.

Niels Boissonnet, Alexis Ghersengorin, and Simon Gleyze. Revealed deliberate preference change. *Games and Economic Behavior*, 142:357–367, 2023. doi: 10.1016/j.geb.2023.08.015.

Pedro Bordalo, Nicola Gennaioli, and Andrei Shleifer. Salience and consumer choice. *Journal of Political Economy*, 121(5):803–843, 2013. doi: 10.1086/673885.

David Broockman and Joshua Kalla. Durably reducing transphobia: A field experiment on door-to-door canvassing. *Science*, 352(6282):220–224, 2016. doi: 10.1126/science.aad9713.

Ian Crawford and Krishna Pendakur. How many types are there? *The Economic Journal*, 123(567):77–95, 2013. doi: 10.1111/j.1468-0297.2012.02545.x.

Franz Dietrich and Christian List. Where do preferences come from? *International Journal of Game Theory*, 42(3):613–637, 2013a. doi: 10.1007/s00182-012-0333-y.

Franz Dietrich and Christian List. A reason-based theory of rational choice. *Noûs*, 47(1):104–134, 2013b. doi: 10.1111/j.1468-0068.2011.00840.x.

Franz Dietrich and Christian List. Reason-based rationalization. Working paper, MPRA Paper 51776, 2013c.

Franz Dietrich and Christian List. Reason-based choice and context-dependence: An explanatory framework. *Economics and Philosophy*, 32(2):175–229, 2016. doi: 10.1017/S0266267115000474.

Sinan Ertemel. Directional revision under two-horizon deliberation: A revealed-preference analysis. Working paper, arXiv:2608.23781, 2026.

Matthew Feinberg and Robb Willer. The moral roots of environmental attitudes. *Psychological Science*, 24(1):56–62, 2013. doi: 10.1177/0956797612449177.

Harry G. Frankfurt. Freedom of the will and the concept of a person. *The Journal of Philosophy*, 68(1):5–20, 1971. doi: 10.2307/2024717.

Till Grüne-Yanoff and Sven Ove Hansson, editors. *Preference Change: Approaches from Philosophy, Economics and Psychology*. Springer, 2009. doi: 10.1007/978-90-481-2593-7.

Takashi Hayashi. Meta-preference, endogenous preference formation and dynamic choice. *Journal of Economic Theory*, 234:106184, 2026. doi: 10.1016/j.jet.2026.106184.

Richard C. Jeffrey. Preference among preferences. *The Journal of Philosophy*, 71(13): 377–391, 1974. doi: 10.2307/2025160.

Botond Kőszegi and Adam Szeidl. A model of focusing in economic choice. *The Quarterly Journal of Economics*, 128(1):53–104, 2013. doi: 10.1093/qje/qjs049.

Kelvin J. Lancaster. A new approach to consumer theory. *Journal of Political Economy*, 74(2):132–157, 1966. doi: 10.1086/259131.

Yusufcan Masatlioglu, Daisuke Nakajima, and Erkut Y. Ozbay. Revealed attention. *American Economic Review*, 102(5):2183–2205, 2012. doi: 10.1257/aer.102.5.2183.

Marcus Pivato. Autonomy and metapreferences. Working paper, SSRN 5410003, 2025a.

Marcus Pivato. Universal recursive preference structures. Working paper, SSRN 5410051, 2025b.

Marcel K. Richter. Revealed preference theory. *Econometrica*, 34(3):635–645, 1966. doi: 10.2307/1909773.

Michael J. Sandel. *The Tyranny of Merit: What's Become of the Common Good?* Farrar, Straus and Giroux, New York, 2020.

Jean R. Soderlund. *Quakers and Slavery: A Divided Spirit*. Princeton University Press, Princeton, NJ, 1985.

Kotaro Suzumura. Remarks on the theory of collective choice. *Economica*, 43(172): 381–390, 1976. doi: 10.2307/2553273.

Edward Szpilrajn. Sur l'extension de l'ordre partiel. *Fundamenta Mathematicae*, 16: 386–389, 1930. doi: 10.4064/fm-16-1-386-389.

Amos Tversky. Elimination by aspects: A theory of choice. *Psychological Review*, 79 (4):281–299, 1972. doi: 10.1037/h0032955.