Essays on

The Behavioral Foundations of Choice

Inauguraldissertation
zur
Erlangung des Doktorgrades
der

Wirtschafts- und Sozialwissentschaftlichen Fakultät
der
Universität zu Köln

2021

vorgelegt
von
Jesper Armouti-Hansen, M.Sc.
aus
Herlev

Referent:	Prof. Dr. Dirk Sliwka
Korreferent:	Prof. Marco Mariotti, Ph.D.
Tag der Promotion:	21.12.2021

Til min familie

Acknowledgements

Firstly, I would like to thank Dirk Sliwka, my supervisor. He has been very generous with his time and always available to discuss new research ideas. Furthermore, his intrinsic motivation for research, expertise and curiosity have been a great source of inspiration. Under his supervision, I have enjoyed great freedom in pursuing my own research ideas.

Next, I would like to express my gratitude to my co-author Christopher Kops for taking on a mentoring role starting all the way back at my time at JGU Mainz, and for his invaluable help and support since then. I would also like to thank Abhinash Borah who, with his inspiring teaching, passion for research and willingness to help students progress and thrive, is one of the main reasons that I have pursued this path.

I would also like to thank former and current colleagues, including co-authors of projects not included in this dissertation, for a pleasant cooperation and numerous interesting discussions during lunch. Many thanks also go to my family and friends for their support.

Lastly, my biggest gratitude goes to my wife, Irini. Firstly, for her unconditional support and for making me belief that I can do things I would never have thought possible. Secondly, for her numerous encouragements during the tougher bits of this process. Finally, for giving me the time and space needed to finish this dissertation by covering for me and being there for our son, Elias. Whatever this work amounts to, I dedicate it to the two of you.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Microeconomic theories of individual behavior are based on choices. That is, individual choices are taken as the primitive concept (i.e., the domain of observables) from which we may postulate models of behavior incorporating preferences, beliefs and potentially other factors that may enter the individual's decision making procedure. Traditionally, the standard framework has been to model individual choice as the result of the maximization of well-defined and stable preferences. Even if this assumption did not always hold, the procedural aspect could be ignored as long as the revealed preferences (i.e., the preferences that are revealed through observable choice) did not conflict with the framework's implied consistency on choice behavior.

Motivated by systematic deviations of this neoclassical rationality consistency requirement, a large amount of research has emerged, in which the procedural aspects of choice are modelled by considering the psychological interpretations of said deviations. Such models, for instance, include non-standard preferences, as the two-stage procedures proposed by Manzini and Mariotti (2007, 2012), or the inclusion of a reference point to which the individual compares a given outcome (Kahneman and Tversky, 1979; Köszegi and Rabin, 2006). A commonality that many such models share is that they are formulated in a mathematical framework that allows us to investigate the intrinsic soundness of its predictions on the domain of choice. In general, this is what this dissertation is about.

One way of conducting this investigation is by providing a behavioral foundation (or, alternatively stated, axiomatic characterization) of the model, which, essentially, is a set of conditions imposed on observable choice behavior, that hold if and only if individuals are choosing according to the model. In addition to allowing for investigating the soundness of its predictions, such a characterization also renders the model falsifiable. An alternative way of doing so, if the proposed model allows for it, is to directly investigate how well the model predicts choices on the considered domain compared to how well it could have predicted. Such an approach additionally informs us how well an alternatively formulated theory could predict. In what follows, the chapters of the dissertation are briefly introduced and summarized. These chapters are structured like regular journal articles, each with its own bibliography. My contribution to each of the chapters can be found in the appendix to the dissertation.

Chapter 2 presents a joint research project with Christopher Kops that is published in Theory and Decision (Armouti-Hansen and Kops, 2018). Herein we propose generalizations of two well-known boundedly rational choice procedures, the rational shortlist method (Manzini and Mariotti, 2007) and the categorize then choose procedure (Manzini and Mariotti, 2012). Our generalization consists of defining these procedures as choice correspondences, instead of choice functions. In turn, this imposes less of a restriction on the contained rationales and allows for the decision maker (DM) to be indecisive as the selection no longer needs to be unique.

Specifically, we consider a DM that chooses by sequentially eliminating inferior alternatives. The method in which alternatives are eliminated is based on pre-defined stage-specific criteria. In the original procedures, a unique alternative remains at the end of this procedure. In our generalization, we only require that the set of alternatives that remain is nonempty. The motivation for doing so is to incorporate indecisiveness. In particular, such a sequential elimination procedure may leave a conflict between the alternatives that is hard to resolve given that more than one alternative remains. If the DM is required to only choose one of these remaining alternatives, the observable choice behavior may be such that we would observe the DM choosing any of these. Hence, if we observe her choice from the same problem multiple times, this may involve different choices.

We provide the axiomatic characterizations of our generalizations by extending the axioms used to characterize the original models. Furthermore, we discuss ways in which an indecisive DM may arrive at a unique choice. In addition, we show that the proposed generalized models can explain behavioral anomalies that cannot be explained in the original setup. This is due to the fact that these anomalies arise when more than one alternative remains after the sequential elimination.

Chapter 3 is also a joint project with Christopher Kops. In it we develop a two-period model of individual decision-making under risk based on recent evidence from neurobiological studies showing that anticipating future outcomes may produce pleasure and pain in the present (Berns et al., 2006, 2007; Schmitz and Grillon, 2012) and research showing that anticipation, in turn, affects the joy from actual consumption (Abeler et al., 2011; Baillon et al., 2020). Our formulated model is such that the DM derives utility from both standard future consumption and non-standard present anticipation of such consumption. In particular, we consider a setting in which the DM may choose her anticipation and where this choice of anticipation, in turn, determines her reference point.

To elaborate on this, consider a DM purchasing a ticket to the state lottery. Quite likely, the DM savors the possibility of becoming rich to the degree that she may become slightly upset when she looses. Analogously, a home owner who fails to invest in home insurance for the upcoming hurricane season may dread a potential disaster until the end of the season brings her more than a relief.

Following the framework of Köszegi and Rabin (2006), we formulate equilibrium concepts that dictates feasible choices of anticipation and consumption lotteries based on when the DM commits to her decision. One requirement that all concepts share is an individual rationality constraint that restricts the DM from anticipating outcomes that are not possible according to her choice of consumption lottery. For instance, if the DM does not purchase a ticket for the state lottery, as in our prior example, she can not anticipate winning the grand prize. In addition, when the DM can only commit to her decision after anticipation, our equilibrium concept requires that, given her choice of consumption lottery, there should be nothing that she can anticipate from this choice that would make her want to choose something else.

We show that our proposed model of anticipation-based and reference-dependent preferences on the domain of choice is equivalent to a two-stage choice procedure. In the first stage of this procedure, a subset of the available alternatives is chosen for consideration based on a filtering process that satisfies well-known internal consistency conditions. In the second stage, the DM applies her preference relation to select the most preferred of the considered alternatives. Furthermore, we provide the axiomatic characterization of this procedure, and hence by extension, a characterization of our model of anticipation-based and reference-dependent preferences. Finally, we show the extent to which the DM's preference and consideration set can be identified from choice data.

Chapter 4 is an empirical study that aims at evaluating how well simple models that incorporate social preferences predict individuals' choices by using machine learning (ML) models as a benchmark. Specifically, we consider panel data from the lab containing experimental observations of binary dictator games and reciprocity games from Bruhin et al. (2019). To evaluate a given model's predictive capability we apply the concept of a model's *completeness* introduced by Fudenberg et al. (2021), which reveals (i) how large a fraction of the predictable variation of the data the model captures, and (ii) how large a gain in performance the model brings compared to a naive baseline model.

To elaborate on this, we define our naive baseline model as one in which the DM only cares about her own payoff. The social preference models that we consider are then designed in a way that sequentially increases the complexity by adding more other-regarding motives. As such, our end point is a linear social preference model that includes potential inequity aversion (or, alternatively, differentiated altruism) and both negative and positive reciprocity.

Our findings on the aggregate level show that the full linear model that includes all of the considered behavioral motives achieves a relatively high completeness estimate of approximately 82%. Thus, the potential improvements of considering more complex functional forms are quite limited on this domain.

We subsequently extend the setting by allowing for the existence of several types. To evaluate the completeness of a model in this setting, we propose two extensions of the original definition. The first proposal *within-type completeness* estimates the completeness within each type that the given model proposes. Hence, this allows us to infer (i) whether there is substantial variation in a model's predictive capability across the types, and (ii) whether a more complex social preference model is needed to fully capture the behavior of some of the types. The second proposal *unrestricted completeness* estimates the completeness of a model with several types by comparing its predictive capability to that of a fully flexible ML model that uses the subject identifier as a feature.

Our within-type completeness results suggest the existence of three types in all of the considered models, with two relatively large ones and one minority type. The choices of the first-type individuals are easily predicted by linear social preference models based on completeness estimates ranging between 88% and 93%. The choices of the second-type individuals, however, seem to be driven by more complex social preference models as we only achieve completeness estimates between 60% and 65% for this type. Finally, due to the relative small size of the minority type, we are unable to properly estimate the within-type completeness of this type.

Finally, the unrestricted completeness estimates suggest that linear social preference models with a parsimonious representation of individuals in the form of three types seems to capture most of the individual variation in data. This is based on completeness estimates between approximately 85% and 88%. However, we stress that these estimates should be seen as upper bounds as there may exist more complex ML methods that lead to better predictions.

1.1 Bibliography

Abeler, J., Falk, A., Goette, L., and Huffman, D. (2011). Reference points and effort provision. *American Economic Review*, 101(2):470–492.

Armouti-Hansen, J. and Kops, C. (2018). This or that? Sequential rationalization of indecisive choice behavior. *Theory and Decision*, 84(4):507–524.

Baillon, A., Bleichrodt, H., and Spinu, V. (2020). Searching for the reference point. *Management Science*, 66(1):93–112.

Berns, G. S., Chappelow, J., Cekic, M., Zink, C. F., Pagnoni, G., and Martin-Skurski, M. E. (2006). Neurobiological substrates of dread. *Science*, 312(5774):754–758.

Berns, G. S., Laibson, D., and Loewenstein, G. (2007). Intertemporal choice–toward an integrative framework. *Trends in Cognitive Sciences*, 11(11):482–488.

Bruhin, A., Fehr, E., and Schunk, D. (2019). The many faces of human sociality: Uncovering the distribution and stability of social preferences. *Journal of the European Economic Association*, 17(4):1025–1069.

Fudenberg, D., Kleinberg, J., Liang, A., and Mullainathan, S. (2021). Measuring the completeness of economic models. Working paper, MIT Economics.

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–292.

Köszegi, B. and Rabin, M. (2006). A model of reference-dependent preferences. *The Quarterly Journal of Economics*, 121(4):1133–1165.

Manzini, P. and Mariotti, M. (2007). Sequentially rationalizable choice. *American Economic Review*, 97(5):1824 – 1839.

Manzini, P. and Mariotti, M. (2012). Categorize then choose: Boundedly rational choice and welfare. *Journal of the European Economic Association*, 10(5):1141–1165.

Schmitz, A. and Grillon, C. (2012). Assessing fear and anxiety in humans using the threat of predictable and unpredictable aversive events (the npu-threat test). *Nature Protocols*, 7(3):527–532.

Chapter 2

This or that? - Sequential rationalization of indecisive choice behavior

This chapter is based on joint work with Christopher Kops.

Abstract. Decision makers frequently struggle to base their choices on an exhaustive evaluation of all options at stake. This is particularly so when the choice problem at hand is complex, because the available alternatives are hard (if not impossible) to compare. Rather than striving to choose the most valuable alternative, in such situations decision makers often settle for the choice of an alternative which is not inferior to any other available alternative instead. In this paper, we extend two established models of boundedly rational choice, the categorize then choose heuristic and the rational shortlist method, to incorporate this kind of "indecisive" choice behavior. We study some properties of these extensions and provide full behavioral characterizations.

Keywords: bounded rationality, choice correspondence, rational shortlist method, categorize then choose, indecisiveness.

JEL Classification Numbers: D01

2.1 Introduction

It is Friday and Herbert has a date for the night. He ponders the question of where to go for dinner. Knowing thyself, he wants to avoid a lengthy and detailed pairwise comparison of all dishes served across the city. Rather, he decides to directly compare entire categories of dinner options (pasta, tacos, tapas) with respect to their cuisine type (Italian, Mexican, Spanish). His experience tells him that he appreciates the Spanish cuisine more than the Italian or the Mexican one, because it is by far his favorite type of cuisine. Browsing through the online menus of Spanish restaurants he stumbles across a small selection of his favorite tapas offered at some of these restaurants and concludes that he cannot narrow down the set of available alternatives any further than to all restaurants serving such food. Our paper builds on this theme of "indecisive" choice behavior and extends previous research on sequential rationalization by Manzini and Mariotti (2007) and Manzini and Mariotti (2012).

Our goal is to explicitly model the procedure of sequential elimination ascribed to Herbert above. Specifically, we consider a decision maker (DM) who chooses according to the following process of sequential elimination: At each stage of the elimination sequence, the DM separately or jointly removes alternatives from further consideration provided that he judges them to be inferior to other available alternatives with respect to certain stage-specific decision criteria just as Herbert uses cuisine type and tapas selection in the example above. At the last stage of such a sequence, the elimination of less attractive alternatives may leave a conflict between the remaining alternatives that is hard to resolve Shafir et al. (1993) to the extent that the DM settles for the choice of some of the remaining alternatives instead of further pursuing the search for the most valuable alternative. This mirrors Herbert's decision to be fine with any restaurant that offers a selection of his favorite tapas.

The literature is rather agnostic about how a DM's conflict between "remaining" alternatives should be interpreted. Eliaz and Ok (2006) suggest that the DM may either be indifferent Kreps (1988) or indecisive Sen (1993) between such alternatives. According to the choice process described above, the remaining alternatives are incomparable for the DM. That is, any remaining alternative is undominated by all other (initially) available alternatives and with respect to any decision criterion used in the elimination sequence. Viewed from this perspective, the term indecisiveness best describes the DM's attitude towards these alternatives. Indeed, under the choice procedures in this paper, it is conceivable that x and y each remain after sequential eliminating alternatives from the set $\{x, y\}$, but sequentially eliminating alternatives from $\{x, y, z\}$ leaves x as the only remaining alternative. If the DM were indifferent between x and y, then, in any choice problem where they are not available, we would expect him to settle for one of the two alternatives if and only if he also considers settling for the other one.

It is this property of indecisiveness that our generalizations of the rational shortlist method (RSM) by Manzini and Mariotti (2007) and the categorize then choose (CTC) procedure by Manzini and Mariotti (2012) distinguish from the respective original versions. In the original formulations, a DM is always able to both identify and pick a unique best alternative, which implies that any indecisiveness has to be resolved across the corresponding sequence of elimination stages. This requires that from all alternatives that remain after the first stage of elimination, the asymmetric binary relation (rationale), which is applied at the second stage (in both models) to remove inferior alternatives, spares exactly one unique maximal element. In our restaurant example, this entails that according to the rationale at the final stage all but one of the dinner options that survive the first stage are dominated by another alternative such that, for instance, there exists only a single undominated tapa that is exclusively offered at one restaurant rather than several favorite tapas offered at different restaurants. Furthermore, the decisiveness demands that no other alternative is chosen if the same choice problem is to be faced repeatedly.

The original version of the CTC by Manzini and Mariotti (2012) is fully characterized by a weak version of the weak axiom of revealed preferences (WWARP), and the characterization of the RSM by Manzini and Mariotti (2007) requires, in addition to this condition, a standard expansion axiom. The authors further note that "in general, we still lack such conditions for general choice correspondences" (Manzini and Mariotti, 2007, p. 1833), so our characterization can also be interpreted as filling this gap. In it, we attempt to closely follow the original axiomatizations by directly transforming these axioms from the domain of choice functions to that of choice correspondences.

Our adjusted version of WWARP keeps the general intuition of the original condition for choice functions, which is that of excluding a certain kind of choice reversals. In our interpretation, choice behavior reveals such a choice reversal if an alternative is chosen over another alternative in some set, but this relation is reversed in a superset of this set. The structure that our axiomatization imposes on choice behavior generally allows for such choice reversals to occur, but rules out choice re-reversals. In other words, our version of WWARP requires that if the same alternative is chosen in a binary comparison with some other alternative and from a set comprising either of these alternatives, then the other alternative can neither be chosen from the set itself nor from any of its subsets that comprise either alternative. Stated differently, our axiom excludes that choice reversals between two alternatives can be reversed again. Our second axiom, a transformation of the original expansion axiom to the structure of choice correspondences, shares its rather straightforward intuition, as it demands that any alternative that is part of the chosen subset of each of two sets is also part of the chosen subset of the union of these two sets.

In Section 2.2 we introduce the setup and formally define our choice procedures. Section 2.3 gleans some intuition about their general properties and Section 2.4 provides their axiomatic characterizations. The final section illustrates a peculiar feature concerning indecisive choice behavior and relates our choice procedures to other models in the axiomatic choice theory literature

2.2 Setup

Let X be a finite set of alternatives, with $|X| > 2$, and let $\mathcal{P}(X)$ denote the set of all nonempty subsets of X. A choice function γ on X selects an alternative from each possible element of $\mathcal{P}(X)$, so $\gamma : \mathcal{P}(X) \to X$ with $\gamma(S) \in S$ for all $S \in \mathcal{P}(X)$. A choice correspondence C_c on X is a mapping C_c: $\mathcal{P}(X) \to \mathcal{P}(X)$ that assigns a subset to each set such that $C_c(S) \subseteq S$ for all $S \in \mathcal{P}(X)$. Given $S \subseteq X$ and an asymmetric binary relation (rationale) $P \subseteq X \times X$, we define the set of P-maximal elements of S as

$$\max(S; P) = \{x \in S | \nexists y \in S \text{ for which } (y, x) \in P\}.$$

Given $S \subseteq X$ and an asymmetric (shading) relation $\succ \subseteq \mathcal{P}(X) \times \mathcal{P}(X)$, denote the set of $\succ$-maximal elements of S by

$$\max(S; \succ) = \{x \in S | \nexists R', R'' \subseteq S \text{ such that } (R', R'') \in \succ \text{ and } x \in R''\}.$$

We abuse notation in a standard way by suppressing set delimiters such that we write $\gamma(x, y)$ instead of $\gamma(\{x, y\})$ and $C_c(x, y)$ instead of $C_c(\{x, y\})$.

The first generalized concept we introduce is the following.

Definition 2.1 *A choice correspondence C_c is an rational shortlist method (RSM) whenever there exists an ordered pair (P_1, P_2) of asymmetric relations, with $P_i \subseteq X \times X$ for $i = 1, 2$ such that:*

$$C_c(S) = \max(\max(S; P_1); P_2) \text{ for all } S \in \mathcal{P}(X)$$

In that case we say that $(P_1; P_2)$ sequentially rationalize C_c.

The choice from any set S can be represented as if the DM sequentially eliminates all other alternatives in two stages. At the first stage, she eliminates all the alternatives that are not maximal according to the first rationale P_1, and from the remaining alternatives she retains, after the second stage, only those specific alternatives that are also maximal according to the second rationale P_2. Any of these remaining alternatives may constitute the DM's choice. In particular, these alternatives are all those that the DM chooses if she repeatedly faces the same choice problem. By this definition the rationales and the order in which they are applied remain the same throughout all choice problems. In general, each relation of the procedure may be incomplete, because the second rationale is not required to be decisive on the alternatives that survive the first stage of elimination.

The following example establishes that the DM's attitude towards the remaining alternatives should be described as indecisiveness, rather than as indifference. It picks up on the discussion raised in the Introduction that if the DM were indifferent between two alternatives, say x and y, then, provided that they are both available, we would expect him to settle for x as his choice if and only if he also considers y to be eligible.

Example 2.2.1 *Let $X = \{x, y, z\}$ be the set of alternatives and $P_1 = \{(z, y)\}$, $P_2 = \{(x, z)\}$ be the DM's rationales. Then, $\max(\max(\{x, y\}; P_1); P_2) = \{x, y\}$ and, thus, by the RSM, $C_c(\{x, y\}) = \{x, y\}$. On the other hand, $\max(X; P_1) = \{x, z\}$ and $\max(\{x, z\}; P_2) = \{x\}$. So, by the RSM, the DM's choice correspondence is $C_c(X) = \{x\}$.*

The second generalized concept we introduce is the following.

Definition 2.2 *A choice correspondence C_c is a categorize then choose (CTC) procedure whenever there exists an asymmetric shading relation $\succ$ on $\mathcal{P}(X)$, with $\succ \subseteq \mathcal{P}(X) \times \mathcal{P}(X)$ and an asymmetric binary relation P on X, with $P \subseteq X \times X$ such that:*

$$C_c(S) = \max(\max(S; \succ); P) \ \textit{ for all } S \in \mathcal{P}(X)$$

In that case we say that $\succ$ and P sequentially rationalize C_c.

As for the RSM, the CTC choice from any set S can be represented as if the DM sequentially eliminates all other alternatives in two stages. At the first stage, she eliminates all categories of alternatives that are not maximal according to the shading relation $\succ$, and from the remaining alternatives she retains, after the second stage, only those specific alternatives that are maximal according to the rationale P. Any of these alternatives is acceptable or satisfactory and might constitute the DM's choice. In particular, these alternatives are all those that the DM chooses if she repeatedly faces the same choice problem. By this definition the shading relation, the rationale and the sequence in which they are applied remain the same throughout all choice problems. In general, each relation of the procedure may be incomplete, because the rationale at the second stage is not required to be decisive on the alternatives that survive the first stage of elimination.

The following example extends Example 2.2.1 to the CTC.

Example 2.2.2 *Let $X = \{x, y, z\}$ be the set of alternatives and $\{x, z\} \succ \{y\}$, $P = \{(x, z)\}$ be the DM's rationales. Then,* $\max(\max(\{x, y\}; \succ); P) = \{x, y\}$ *and, thus, by the CTC, $C_c(\{x, y\}) = \{x, y\}$. On the other hand, it holds that*$\max(X; \succ) = \{x, z\}$ *and* $\max(\{x, z\}; P) = \{x\}$*. So, by the CTC, the DM's choice correspondence is $C_c(X) = \{x\}$.*

Remark 2.1 *In both definitions above P is an asymmetric binary relation in the form of a strict rationale that captures the DM's indecisiveness between any two remaining alternatives. These definitions provide the most natural extensions of their original counterparts. On the other hand, it is possible to introduce indifference into the models by defining P to be a weak rationale. In this case, however, it is not clear whether the DM is indecisive or indifferent between any two "weakly" P-maximal elements of S. Indeed, it may well be the case that she is indifferent between some of the remaining alternatives and indecisive between others.*

2.3 An example

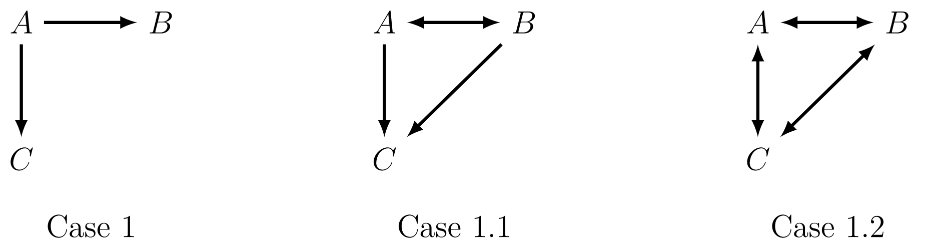
Consider a DM who frequently invests in the stock market and assume that she chooses among three different kinds of shares A, B and C. We assume that common fluctuations in supply and demand can induce temporary unavailability of each of these assets such that we are able to observe choices from the grand set $\{A, B, C\}$ as well as from any of its nonempty subsets. For our example the intuition of a choice correspondence is as follows: when facing a certain set of available shares, the DM may select any of its nonempty subsets. If the DM selects a subset that comprises just a single asset, she will also choose this item. If the selected subset is not a singleton, the shares in that set are those that the DM might choose, i.e., those are her acceptable alternatives.[1]

For each of the cases that follow below we fix the single item that is always part of the selected subset of shares, singleton or not, when all three different shares are available, to asset A. Anytime choices are decisive such that the DM is able to identify and pick a single best share the generalized version of the choice procedures that we consider here coincides with their respective original version. This applies to Case 1, 2, and 3, whereas the other cases are specially geared to choice correspondences. Cases 1, 1.1 and 1.2 treat instances of (in)decisive rational choice and the choice cycle in Case 2 is extended to indecisive choice behavior in Case 2.1 and Case 2.2. The choice behaviors illustrated in Case 3.1 and Case 3.2 show aspects of indecisive choice that cannot be captured in the decisive counterpart of default choice in Case 3. Finally, Case 4 and Case 4.1 highlight a peculiar feature of indecisive choice behavior which we refer to as reversed Condorcet inconsistency. The arrows in the following figures point away from the alternative that is (or may be) chosen in pairwise comparisons, the double arrow indicates that both alternatives are acceptable, i.e., may be chosen.

Figure 2.1 (dominance of the best share(s)): in Case 1, the DM chooses asset A whenever it is available, regardless of her choice when it is not available. In Case 1.1, the DM selects the subset $\{A, B\}$, when both shares are available, but she never chooses share C when any other item is available. In Case 1.2, the DM always selects the entire

[1] Rubinstein and Salant (2006) provide another interpretation of choice correspondences that results from choice that is sensitive to the order in which the decision maker is confronted with the available alternatives. Namely, choice correspondences, which attach to every set of alternatives all the elements that are chosen for some ordering of that set.

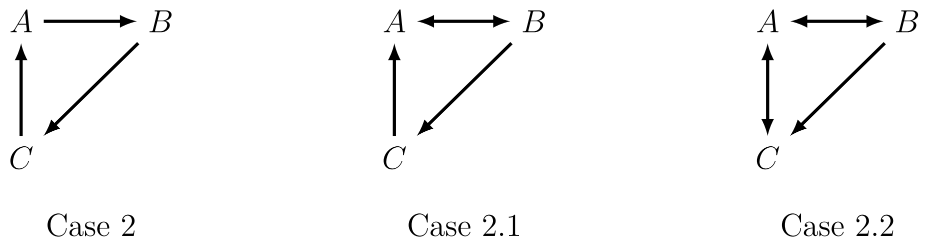
Figure 2.1: Dominance of the best share(s)



Case 1 Case 1.1 Case 1.2

available set. If we let asset A be the single best item in 1, A and B be the best assets in 1.1, and all three shares be equally good in 1.2, the choice behavior described above for each of these scenarios can be represented by the maximization of an asymmetric and negatively transitive order,[2] i.e., it does not violate WARP for choice correspondences.

Figure 2.2: Pairwise cycle of choice



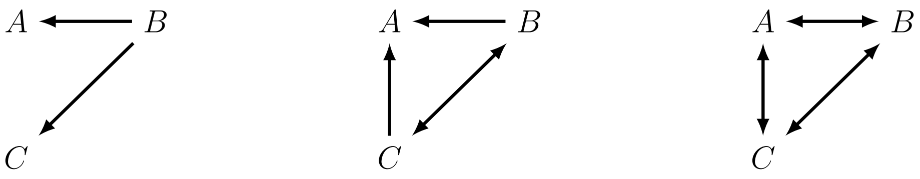
Case 2 Case 2.1 Case 2.2

Figure 2.2 (pairwise cycle of choice): in Case 2, the DM chooses asset A from the grand set and when B is the only other available share, B when C is the only other available share, and C when A is the only other available share. In Case 2.1, the DM's choices among $\{A, C\}$ and $\{B, C\}$ remain the same as in 2, i.e., A in the former and B in the latter, but the DM now selects the entire set $\{A, B\}$, whenever it is available. In Case 2.2, the DM's choices among $\{A, B\}$ and $\{B, C\}$ remain the same as in 2.1, but the DM now also selects the entire set $\{A, C\}$, when it is available. Clearly, there does not exist a single asymmetric and negatively transitive order that can explain any of these three cases, i.e., such choice behavior violates WARP. However, we can rationalize this choice behavior by the sequential application of two rationales. Let us call P_1 popularity and P_2 proximity and let the DM prefer the shares of more popular companies to those of less popular ones and shares of domestic companies to foreign ones. Further, let the DM know that share B is more popular than share C, but suppose that she is unable to rank company A relative to B and C on this parameter. That is, she cannot judge

[2] A binary relation $\succ$ is negatively transitive on X if $x \not\succ y$ and $y \not\succ z$ implies $x \not\succ z$ for all $x, y, z \in X$.

whether company A is more or less popular than B and C, respectively. Let C be issued by a company that is closer to the DM's place of abode than the issuer of A and lastly B as the share of the company with the farthermost headquarters. Let the DM look first at the popularity and then at proximity, the choices in 2 can then be rationalized by applying the rationales in this order. Suppose now that the issuers of share A and B are indistinguishable in proximity but that the rest remains the same as before, then the choices in 2.1 can again be rationalized by applying the criteria popularity and proximity in that order. Finally, suppose that in 2.2 all share issuers are incomparable with regard to their proximity, then the choices in 2.2 can be rationalized by applying the criteria popularity and proximity in that order.

Figure 2.3: Default share(s)

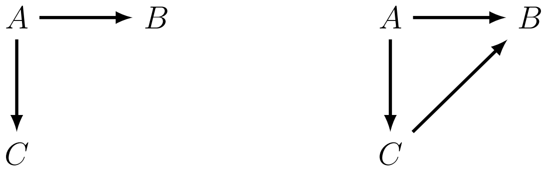


Case 3 Case 3.1 Case 3.2

Figure 2.3 (default share(s)): fix the choice behavior such that only A is chosen from the grand set, and, in all three cases of binary choice, B is part of the selected subset whenever it is available. Clearly, none of the three cases can be represented by a single asymmetric and negatively transitive order on the grand set as the choice behavior in each case violates WARP. Furthermore, such choice behavior cannot be rationalized by the RSM. To see this, suppose by contradiction that it was the RSM with the rationales popularity and proximity. Since in all three cases share B is always part of the selected subset when it is available in pairwise comparisons with A and C, neither A nor C can be more popular than B. If A (or C) survives the first stage, then its issuing company cannot be closer to the DM's place of abode than the issuer B as B is chosen in all pairwise comparisons. This implies that B can never be eliminated by a sequential application of the popularity and proximity rationales, which is a contradiction to B not being chosen from the grand set. So, there exists no RSM that can represent any choice behavior in Figure 2.3.

The CTC procedure, in turn, can rationalize either choice behavior in Figure 2.3. To see this, let the shading relation be popularity and let the DM know that share A is the most popular asset when all three shares are available, but if only two shares are available then popularity is uniformly distributed across the available shares. This implies that share A is the unique choice from the grand set. Further, let the binary relation that is applied after the shading stage be proximity. Let B be issued by the only domestic company in Case 3, let the issuer of B and C be both domestic and that of A be foreign in Case 3.1 and let all assets be domestic ones in Case 3.2. Then each case can be rationalized by the respective ensuing CTC procedure with the shading relation of popularity and the rationale of proximity.

Figure 2.4: Reversed Condorcet inconsistency



Case 4 Case 4.1

Figure 2.4 (reversed Condorcet inconsistency): for each case, fix $\{A, B\}$ to be the subset that is selected from the grand set. In Case 4 and in Case 4.1, asset A is part of the selected subset whenever it is available in pairwise comparisons and share B is not chosen from the pairwise comparison with A.

Clearly, any choice behavior here violates WARP, so there does not exist a single asymmetric and negatively transitive order that can represent the choice data. Further, these choices can neither be rationalized by the RSM nor by the CTC. Note that any RSM is a special case of a CTC, so to prove this statement it suffices to show that choices cannot be rationalized by any CTC. Suppose by contradiction that there exists a generalized CTC that can represent these choices, w.l.o.g. this CTC has popularity as the shading relation and proximity as the binary relation at the second stage. First, neither A nor B can be eliminated with respect to popularity in the grand set, because this contradicts $\{A, B\}$ being selected from that set. In particular, asset C can neither be more popular nor be issued by a company that is closer to the DM's place of abode than the issuer of A or B, considering that $\{A, B\}$ is selected from the grand set. Second, the issuer of B has to be relatively more distant than that of A given that A is the unique choice from the binary comparison with B. This implies that in the grand set, A eliminates B by the proximity rationale, which contradicts $\{A, B\}$ being selected from the grand set. Hence, choice behavior in Case 4 and Case 4.1 cannot be rationalized by any CTC.

2.4 Characterizations

2.4.1 Rational shortlist method

The first property we introduce for the characterization of the generalized RSM captures the intuition of choice reversals. In contrast to standard rational choice, the choice procedure of the RSM explicitly allows for choice reversals to occur, but not for choice re-reversals as we explained in the Introduction. This is formalized in the following axiom.

Axiom 2.1 (Weak WARP*) *If x is the unique choice in a binary comparison with y and x is chosen when y and other alternatives $\{z_1, \ldots, z_k\}$ are available, then y is not chosen when x and a subset of $\{z_1, \ldots, z_k\}$ are available. Formally, for all S, $T \in \mathcal{P}(X)$:*

$$[\{x, y\} \subset S \subseteq T, y \notin C_c(x, y) \text{ and } x \in C_c(T)] \Rightarrow [y \notin C_c(S)]$$

The second property is a standard expansion axiom.

Axiom 2.2 (Expansion*) *An alternative chosen from each of two sets is also chosen from their union. Formally, for all S, $T \in \mathcal{P}(X)$:*

$$[x \in C_c(S) \cap C_c(T)] \Rightarrow [x \in C_c(S \cup T)]$$

The axioms of Weak WARP* (WWARP*) and Expansion* are the only properties we use in our characterization, so we can now state our first main result as follows.

Theorem 2.1 *Let X be any finite set. A choice correspondence C_c on X is an RSM, if and only if it satisfies Expansion* and WWARP*.*

Proof: Necessity: Let C_c be an RSM on X and let P_1, P_2 be the rationales.

(a) Expansion*: Let $x \in C_c(S) \cap C_c(T)$, for S, $T \in \mathcal{P}(X)$. For Expansion* to hold we have to show that this implies $x \in C_c(S \cup T)$. For this it is enough to show that for any $y \in S \cup T$, it cannot be $(y, x) \in P_1$ and for any $y \in \max(S \cup T; P_1)$, it cannot be

$(y,x) \in P_2$. Suppose $(y,x) \in P_1$ for some $y \in S \cup T$, then either $y \in S$ or $y \in T$ which would immediately contradict $x \in C_c(S)$ or $x \in C_c(T)$. Hence, this would contradict C_c being an RSM. Suppose, now, that for some $y \in \max(S \cup T; P_1)$ we had $(y,x) \in P_2$. Since $\max(S \cup T; P_1) \subseteq \max(S; P_1) \cup \max(T; P_1)$ we have $y \in \max(S; P_1)$ or $y \in \max(T; P_1)$ contradicting $x \in \max(\max(S; P_1); P_2)$ or $x \in \max(\max(T; P_1); P_2)$. Hence, $x \in C_c(S \cup T)$, so Expansion* holds.

(b) WWARP*: Let $y \notin C_c(x,y)$, $x \in C_c(T)$, $y \in T$. For WWARP* to hold we have show that for any S with $\{x,y\} \subset S \subseteq T$ we have $y \notin C_c(S)$. The fact that $y \notin C_c(x,y)$ implies that $(x,y) \in P_1 \cup P_2$, i.e., either $(x,y) \in P_1$ or P_1 is indecisive and $(x,y) \in P_2$. Suppose $(x,y) \in P_1$, then $y \notin S$ follows immediately. Suppose $(x,y) \in P_2$. The fact that $x \in C_c(T)$ implies that for all $z \in S$ it is the case that $(z,x) \notin P_1$. Therefore, x never drops out by P_1, i.e., $x \in \max(S; P_1)$ for all $S \subseteq T$ for which $x \in S$. Since $(x,y) \in P_2$, then $y \notin \max(\max(S; P_1); P_2)$ for all such S, and thus $y \notin C_c(S)$.

Sufficiency: Suppose that C_c satisfies the axioms, i.e., WWARP* and Expansion*. We construct the rationales explicitly. Define

$$P_1 = \{(x,y) \in X \times X | \nexists S \in \mathcal{P}(X) \text{ such that } y \in C_c(S) \text{ and } x \in S\}$$

, i.e., $(x,y) \in P_1$ if and only if y is never chosen when x is also available for all sets $S \in \mathcal{P}(X)$. Next, define $(x,y) \in P_2$ if and only if $y \notin C_c(x,y)$, i.e., $(x,y) \in P_2$ if and only if y is not chosen in the direct comparison $\{x,y\}$.

By these definitions P_1 and P_2 are both asymmetric: If $(x,y) \in P_1$ and $(y,x) \in P_1$, then $x, y \notin C_c(x,y)$ which is not possible as by definition $C_c(.)$ assigns a nonempty subset. If $(x,y) \in P_2$ and $(y,x) \in P_2$, then $x, y \notin C_c(x,y)$ which is not possible as by definition $C_c(.)$ assigns a nonempty subset.

To check that P_1 and P_2 rationalize C_c, take any $S \in \mathcal{P}$ and let $x \in C_c(S)$. First, we show that all alternatives that are chosen over x in binary choice are eliminated in the first round by P_1. Second, we show that x survives both rounds, i.e., x is neither eliminated by P_1 nor by P_2.

First, let $z \in S$ be such that $x \notin C_c(x,z)$. Suppose, by contradiction, that for all $y \in S \backslash \{z\}$ there exists $T_{yz} \in \mathcal{P}(X)$, $y, z \in T_{yz}$, such that $z \in C_c(T_{yz})$. By Expansion* $z \in C_c(\bigcup_{y \in S \backslash \{z\}} T_{yz})$. If $S = \bigcup_{y \in S \backslash \{z\}} T_{yz}$ we have $z \in C_c(S)$ which together with WWARP* yields $x \notin C_c(S)$, i.e., a contradiction to $x \in C_c(S)$. If $S \subset \bigcup_{y \in S \backslash \{z\}} T_{yz}$, then WWARP* yields again $x \notin C_c(S)$, i.e., a contradiction to $x \in C_c(S)$. Thus for all z with $x \notin C_c(x, z)$ there exists $y_z \in S$ such that $(y_z, z) \in P_1$

Second, x is not eliminated by either P_1 or P_2. For any $y \in S$, if $(y, x) \in P_1$ then by definition of P_1, $\nexists T \in \mathcal{P}(X)$ such that $x \in C_c(T)$, but this contradicts $x \in C_c(S)$. If $(y, x) \in P_2$, then y would be chosen over x in binary choice, i.e., $x \notin C_c(x, y)$, and, by the argument in the previous paragraph, y would have been eliminated by the application of P_1 before P_2 can be applied. □

2.4.2 Categorize then choose

For our second main result the only property we use is that of the axiom of WWARP*, so we can state our second main result as follows.

Theorem 2.2 *Let X be any finite set. A choice correspondence C_c on X is a CTC, if and only if it satisfies WWARP*.*

Proof: Necessity: Let C_c be a CTC on X with a shading relation $\succ$ and a binary relation P. Suppose $y \notin C_c(x, y)$ and $x \in C_c(S)$ for some S with $y \in S$. Now suppose by contradiction that $y \in C_c(R)$ for some R with $x \in R \subseteq S$. This implies that $x \notin \max(R, \succ)$, since $y \notin C_c(x, y)$ implies $(x, y) \in P$ (note that the possibility $\{x\} \succ \{y\}$ yields an immediate contradiction with $y \in C_c(R)$). In particular, there exist $R', R'' \subseteq R$ such that $R' \succ R''$ and $x \in R''$. Since $R', R'' \subseteq S$ this fact contradicts $x \in C_c(S)$.

Suffiency: Suppose that C_c satisfies WWARP*. We construct the rationale and the shading relation explicitly. Define $(x, y) \in P$ if and only if $y \notin C_c(x, y)$. P is clearly asymmetric, but may be incomplete. If $(x, y), (y, x) \in P$, then $x, y \notin C_c(x, y)$ which violates the property $C_c(.) \neq \emptyset$ of the definition of choice correspondences. Next, we establish that P is well-defined. That is, it cannot be that $y, z \in C_c(S)$ for some $S \in \mathcal{P}(X)$ and $(y, z) \in P$. Suppose by contradiction that there exist $S \in \mathcal{P}(X)$ and $y, z \in C_c(S)$ such that $(y, z) \in P$. Then $z \notin C_c(y, z)$ and $y \in C_c(S)$, but $z \in C_c(R)$ for $R = S$, which contradicts the fact that C_c satisfies WWARP*. So, P is well-defined. Fixing the choice correspondence C_c, we define the upper and lower contour sets of an alternative on a set $S \in \mathcal{P}(X)$ as

$$\mathrm{Up}(C_c(S), S) = \{y \in S \backslash C_c(S) | (y, x) \in P \text{ for } x \in C_c(S) \ \vee \ (x, y) \notin P, \ \forall x \in C_c(S)\}$$

and

$$\text{Lo}(C_c(S), S) = \{y \in S | (y, x) \notin P, \forall x \in C_c(S) \ \wedge \ (x, y) \in P \text{ for } x \in C_c(S)\}$$

respectively. Define $R \succ S$ if and only if there exists a $T \in \mathcal{P}(X)$ such that

$$R = C_c(T) \cup \text{Lo}(C_c(T), T)$$

and

$$S = \text{Up}(C_c(T), T) \neq \emptyset$$

By this definition $\succ$ is asymmetric. If $(R', R'') \in \succ$ and $(R'', R') \in \succ$, then $C_c(R' \cup R'')$ $= \emptyset$ which by definition is not possible.

Now, let $S \in \mathcal{P}(X)$ and let $x \in C_c(S)$. Suppose $(y, x) \in P$ for some $y \in S$. Then by construction

$$C_c(S) \cup \text{Lo}(C_c(S), S) \succ \text{Up}(C_c(S), S)$$

and $y \notin \max(S, \succ)$.

Next, suppose by contradiction that $x \notin \max(S, \succ)$. Then there exists $R', R'' \subset S$ with $R' \succ R''$ and $x \in R''$. Define $R = R' \cup R''$. By construction of $\succ$, we must have

$$R' = C_c(R) \cup \text{Lo}(C_c(R), R)$$

and

$$R'' = \text{Up}(C_c(R), R)$$

and $x \notin C_c(R)$. This means that $(x, y) \in P$ for some $y \in C_c(R)$ and, therefore, $y \notin C_c(x, y)$, for this $y \in C_c(R)$. As $R = R' \cup R'' \subseteq S$ and $y \in C_c(R)$, the fact that $y \notin C_c(x, y)$ together with $x \in C_c(S)$ contradicts WWARP*.

Finally, by construction we have that $(x, y) \in P$, for all $y \in \max(S, \succ) \backslash C_c(S)$ and some $x \in C_c(S)$ (since then $y \in \text{Lo}(C_c(S), S)$). □

2.5 Beyond sequential elimination

2.5.1 Indecisiveness vs. indifference

As we have defined our models in Section 2.2, it is, from our perspective, intuitive to think of the remaining alternatives from a set after the sequential elimination as the alternatives among which the DM is indecisive. This is because the alternatives that remain constitute the set of undominated (according to a strict relation), but not dominating (according to a weak relation), options. Thus, each pair of these alternatives is incomparable with respect to the rationale applied in the final stage. Hence, by the definition of the RSM and CTC, if the choice correspondence is a non-singleton set, it must be that the DM is indecisive between these alternatives, as the models allow no scope for indifference.

One could increase the scope of the models to include indifference by allowing the relations used in the sequential elimination to be weak. In the following, we discuss a generalization of the RSM choice procedure, but the same argumentation can easily be applied to the CTC. In particular, we let the interpretation of the reflexive rationale $P' \subseteq X \times X$ be such that $(x, y) \in P'$ means "x is at least as good as y according to P'" and derive the asymmetric strict rationale $P \subseteq P'$ such that $(x, y) \in P$ if and only if $(x, y) \in P'$ and $(y, x) \notin P'$. Notice now that there are two ways in which the DM can maximize according to the weak rationale P':

$$\max(S, P') = \{x \in S | \nexists y \in S \text{ such that } (y, x) \in P' \text{ and } (x, y) \notin P'\} \tag{2.1}$$

and

$$\max(S, P') = \{x \in S | \forall y \in S, (x, y) \in P'\} \tag{2.2}$$

In words, (2.1) states that $x \in S$ is maximal according to P' if, whenever there exists an alternative $y \in S$ which is "at least as good as x according to P'", it must be that the two alternative are in fact equivalent according to P'. Notice, however, that this is equivalent to the maximization of the strict rationale P, i.e.,

$$\max(S, P) = \{x \in S | \nexists y \in S \text{ for which } (y, x) \in P\} \tag{2.3}$$

which is the maximization principle used in the definition of the RSM in Section 2.2. The maximization of S according to P' used in (2.2) states that $x \in S$ is maximal according to P' if it is "at least as good as every $y \in S$ according to P'", that is, if x dominates all other alternatives in S according to P'. It follows immediately that the set of maximal alternatives from (2.1) and (2.2) may not coincide. Whereas (2.2) will be empty whenever (2.1) is empty, the inverse may not be true. In fact, as Danan (2003) states, (2.2) is nonempty on all binary subsets of S if and only if P' is complete. Thus, (2.2) allows no scope for indecisiveness. Hence, for the sake of generality, we consider (2.1) here.[3] Further, if we keep the assumption that the strict relations applied in stage one and two may be jointly incomplete, one may ask oneself: even if the preference judgments applied by the DM are observable to us, can we then distinguish between alternatives among which the DM is indecisive from those among which she is indifferent? For example, suppose that, when choosing between alternative A and B, the DM finds the pair incomparable according to the rationale used in the first stage (i.e., neither $(A, B) \notin P_1'$ nor $(B, A) \notin P_1'$) and, according to the rationale used in the second stage, she finds the alternatives equivalent (i.e., $(A, B), (B, A) \in P_2'$). It would then be intuitively appealing for us to think of the DM being indifferent between A and B. However, if the situation is reversed such that she is indifferent in the first stage and indecisive in the second, we would rather think of her being indecisive between A and B. In the case that the alternatives are incomparable (equivalent) according to the rationales in both stages, we would naturally think of the DM being indecisive (indifferent) between the two. If we use this interpretation, we can answer the question in the affirmative. Nevertheless, the separation between indecisiveness and indifference becomes fluid once introducing multiple criteria for decision making. If we stick to the assumption of the revealed preference approach that we can only observe an individual's choices but not her preference judgments, distinguishing between indecisiveness and indifference is difficult, if not impossible.

In their method of distinguishing between indifferent and indecisive pairs of alternatives, Eliaz and Ok (2006) call a pair of alternatives (x, y) C_c-incomparable if they are both chosen in their binary comparison and if the alternative y is not equivalent to the alternative x in one the following ways: (i) if x is chosen from a set S not including y, then removing x and including y does not lead y to be chosen, (ii) if from the same set S, x is not chosen, then removing x and including y leads y to be chosen, or (iii) if the set of chosen alternatives of these two sets are not the same when excluding x and y. The authors show that if the DM is behaving as if she is maximizing according to a possibly incomplete weak preference relation[4], i.e., $C_c(\cdot) = \max(\cdot, \succsim)$, then a C_c-incomparable pair is a pair between which the DM is indecisive according to the preference relation. How-

[3] Note that our characterization does not change by considering this more general version of the choice procedure.

[4] That is, a reflexive and transitive (but not necessarily complete) binary relation.

ever, since we do not require transitivity, it can easily be shown that this notion will not necessarily reveal indecisiveness in our setting. Danan (2003), on the other hand, uses the link between incompleteness of tastes (i.e., indecisiveness) and preference for flexibility to derive a DM's preferences. In particular, by defining preferences on menus of alternatives and requiring that incompleteness of preferences implies a strict preference for flexibility, Danan (2003) shows that incompleteness of preferences can be uniquely derived from observed choice behavior. Specifically, a DM has a preference for flexibility among the menus $S, T \in \mathcal{P}(X)$ if and only if she strictly prefers the joint menu $S \cup T$ over S and T, respectively. Savage (1954) argues that indifference between two alternatives can be revealed by an experimental test. In particular, he suggests that indifference is indirectly revealed when adding an arbitrarily small monetary bonus to one of the two alternatives changes a decision maker's choices between these two alternatives. Mandler (2009) shows that indifference and indecisiveness cannot be distinguished from observable choice behavior in a standard one-shot choice setting. Furthermore, the author shows that, in a sequential choice setting, even a sequentially rational choice correspondence, i.e., a choice correspondence that selects all undominated alternatives according to a strict preference relation, may lead to an irrational chain of trades leaving the DM strictly worse off. Finally, Hill (2012) tackles the issue of incompleteness of preferences by formally introducing the notion of confidence in preferences. The conjecture is that the more important a choice problem is, the more confident a DM has to be in her preferences to make a choice according to it. The author provides an axiomatic characterization of choice behavior depending on both the issue's importance as well as the DM's confidence.

If the DM is required to choose one, and only one among the alternatives between which she is indecisive, one naturally must discuss ways of which she overcomes this indecisiveness and ends up with a decision. Rubinstein and Salant (2008), for whom a choice correspondence can be seen as a notion capturing indeterminacy of choice, extends the standard choice problem by the inclusion of a frame f. Here, an extended choice problem is captured by the double (S, f), where $S \in \mathcal{P}(X)$ is the standard set of feasible alternatives and $f \in F$ is a frame which may or may not effect the DM's choice. By this definition, an extended choice function $\gamma : \mathcal{P}(X) \times F \to X$ is a mapping that from any extended choice problem $(S, f) \in \mathcal{P}(X) \times F$ chooses a single alternative $x \in S$. The interpretation of this choice function is that the choice may be sensitive to the frame itself. For example, Rubinstein and Salant (2006) consider a choice function from lists, where altering the order in which the alternatives are displayed may alter the choice. Therefore, as stated by Rubinstein and Salant (2008), one can think of the choice correspondence $C_c(S)$ as the set of alternatives chosen by an extended choice functions on S for some frame f. Based on this thought, one may think of our RSM choice correspondence as a correspondence induced by RSM choice functions sensitive to frames. In particular, we can define an extended RSM choice function as

$$\gamma(S, f) = \max(\max(\max(S; P_1); P_2); f) \text{ for all } S \in \mathcal{P}(X), f \in F$$

and require that the frame is able to reduce the choice to a singleton $x \in S$ any time the rationales fail to do so on any $S \in \mathcal{P}(X)$. Returning to the choice function from lists, one could think of a DM choosing the first alternative in a list whenever more than one alternative remains after the sequential elimination. Then, our general RSM can formally be defined as

$$C_c(S) = \bigcup_{f \in F} \{\max(\max(\max(S; P_1); P_2); f)\} \text{ for all } S \in \mathcal{P}(X)$$

implying that the correspondence and thus the characterization give forth the observable choices observable when choice is sensitive to an impalpable frame. Closely related to this is the model of behavioral data sets by Rubinstein and Salant (2012). Here it is assumed that a DM has an underlying preference relation $\succ$, but that there might exist factors, such as frames, affecting her ordering of alternatives. With this conjecture in mind, the authors define a distortion function $D(\succ)$ as a mapping from an ordering to the set of all possible orderings a DM may display conditional on this ordering. Based on this function, one can conclude that the behavioral data set Λ, which is also a set of orderings, revealed by the DM, is consistent with the distortion of a given ordering if $\Lambda \subseteq D(\succ)$. This setup can be adapted to our setting by checking if a behavioral data set Λ is consistent with a distortion of a DM choosing according to the RSM. That is, if $\Lambda \subseteq D(P_1, P_2)$.

2.5.2 Pairwise and Condorcet consistency

We have shown in Section 2.3 that the generalized RSM accommodates instances of pairwise choice cycles and that the generalized CTC model can, in addition to such cyclical choice behavior, also explain a specific kind of choice reversals. Furthermore, we have shown that there exists a third type of such "choice pathologies" that none of these models can rationalize. In this section, we show that it is this pathology and the violation of WARP it induces that the axiomatization of a choice procedure has to address provided that the procedure renders indecisive choice behavior possible. In fact, this pathology does not arise for decisive choice behavior rather it is an exclusive feature of indecisive choice behavior.

For the sake of completeness, we briefly restate the intuition of WARP, that is, if an alternative is chosen over another alternative within a certain set of alternatives, then changing the set cannot reverse this choice behavior. Formally, this axiom can be stated as follows.

Definition 2.3 *WARP: For all S, $T \in \mathcal{P}(X)$*

$$\Big[x = \gamma(S),\ y \in S,\ x \in T\Big] \Rightarrow \Big[y \neq \gamma(T)\Big]$$

For choice functions, we can decompose violations of WARP into two independent choice pathologies which we illustrate in Figure 2.5.

Figure 2.5: Choice pathologies for decisive and indecisive choice behavior

$$\begin{tikzcd}
\{x,y\} \arrow[r] & x & \{x,y\} \arrow[r] & x & \{x,y\} \arrow[r] & x \\
\{y,z\} \arrow[r] & y & \{y,z\} \arrow[r] & y & \{y,z\} & \\
\{x,z\} \arrow[r] & x & \{x,z\} \arrow[r] & z & \{x,z\} \arrow[r] & x \\
\{x,y,z\} \arrow[r] & x & \{x,y,z\} & & \{x,y,z\} \arrow[r] & y
\end{tikzcd}$$

Rational Choice Pathology 1 Pathology 2

In "Pathology 1", the choices exhibit a binary cycle, as x is chosen from $\{x, y\}$ and y is chosen from $\{y, z\}$, but z is chosen from $\{x, z\}$. Choice behavior that reveals such a binary cycle is pairwise inconsistent with rational choice, independent of what alternative is chosen from the set that contains all alternatives of the cycle, i.e., independent of what is chosen from $\{x, y, z\}$. This is so because whatever alternative is chosen from $\{x, y, z\}$, we can remove one of the two unchosen alternatives such that from the resulting set the other remaining unchosen alternative is now chosen over the alternative that is chosen from $\{x, y, z\}$. If, for instance, x is chosen from $\{x, y, z\}$, we can remove y such that from the resulting set $\{x, z\}$, z is chosen over x.

In “Pathology 2”, x is chosen in the respective pairwise comparisons with y and z, but fails to be chosen from $\{x, y, z\}$. The corresponding choice behavior is Condorcet inconsistent with rational choice, because the alternative that is chosen from each pairwise comparison with any alternative of a certain set is not chosen from that set, i.e., x is chosen from $\{x, y\}$ and $\{x, z\}$, but not from $\{x, y, z\}$.

If a choice procedure allows choice behavior to be indecisive, then a third pathology may arise. This pathology takes the general form of the choice behavior illustrated in Figure 2.6.

Figure 2.6: Additional choice pathology for indecisive choice behavior

$$\begin{array}{c}
\{x, y\} \longrightarrow \{y\} \\
\{y, z\} \longrightarrow \{y\} \\
\{x, z\} \longrightarrow \{x\} \\
\{x, y, z\} \longrightarrow \{x, y\}
\end{array}$$

Pathology 3

In “Pathology 3”, x is chosen from $\{x, y, z\}$, but it is not chosen from each pairwise comparison with other alternatives of that set, more precisely, x is not chosen from $\{x, y\}$. This pathology reverses the intuition of the second pathology, so the corresponding choice behavior is Condorcet inconsistent with rational choice in the reversed way, because it pertains to a situation in which an alternative that is chosen from a certain set, x from $\{x, y, z\}$, is not chosen from each pairwise comparison with any alternative of that set, i.e., x is not chosen from $\{x, y\}$.

This type of inconsistency cannot arise for procedures that require choice behavior to be decisive and neither the generalized version of the RSM nor that of the CTC procedure can accommodate this choice pathology. An elaborate explanation of this fact is given after Figure 2.4 in Section 2.3, but it is also immediate from the transformation of the axiom of WWARP to indecisive choice behavior in the previous section. In contrast to WWARP, the axiom of WWARP* also requires that choice re-reversals are excluded across just two sets, i.e., if some alternative is not chosen in a pairwise comparison with another alternative then there exists no set from which both alternatives are selected. This highlights that “Pathology 3” constitutes a violation of WWARP* given that x is not chosen from the pairwise comparison with y, but either alternative is chosen from $\{x, y, z\}$.

2.5.3 Other sequential procedures of choice

In the rational shortlist method, Manzini and Mariotti (2007) consider a sequential choice procedure defined by two asymmetric binary relations where these relations and the sequence in which they are applied are invariant with respect to the choice set. The fact that the definition of the RSM requires choice behavior to be decisive makes it a special case of our corresponding generalization, the RSM*. In a companion paper, Manzini and Mariotti (2012) provide a characterization of a variation of this choice procedure in which the first asymmetric relation in the sequence of rationales is not restricted, by definition, to binary comparisons. The ensuing CTC choice procedure requires choice behavior to be decisive which makes it a special case of our corresponding generalization, the CTC*. Au and Kawai (2011) restrict the RSM model to the use of transitive rationales. The axiom they introduce for this purpose is not affected by indecisiveness of choice behavior. The rationalization model of Cherepanov et al. (2013) generates the same choice data as a CTC choice procedure does such that it is also fully characterized by the axiom of WWARP. Our reformulated version of this axiom, WWARP*, presumably suffices to fully characterize the obvious variation of the rationalization model to indecisive choice behavior.

In the choice procedures described by Kalai et al. (2002) and Apesteguia and Ballester (2005) multiple rationales are used to explain choice behavior. Their focus is not on a sequential application of several rationales, rather the authors are interested in identifying the minimum number of rationales necessary to explain choice data if the application of each relation can be restricted to specific subsets.

A completely different approach to choice behavior is taken by Masatlioglu et al. (2012) in their revealed attention model. According to their two-stage procedure of choice with limited attention (CLA), first an attention filter determines which of the available alternatives are feasible and then the DM selects the unique option from the set of feasible alternatives that maximizes a complete and transitive binary relation. The authors show that their model is fully characterized by a single axiom called Limited Attention WARP (LA-WARP) which shares no logical connection to the characterizations of the RSM and the CTC. Furthermore, they provide examples of an RSM that cannot be a CLA and the other way around which suggests that a transformation of the CLA model to indecisive choice behavior presumably retains this lack of a logical connection to the generalized RSM and CTC.

In a companion paper, Lleras et al. (2014) introduce a variation of the revealed attention model that they refer to as the limited consideration model. Choice with limited consideration (CLC) is a two-stage choice procedure in which, at the first stage, a consideration filter restricts the set of available alternatives to feasible ones and then in the second stage the DM selects the unique option from the set of feasible alternatives that maximizes a complete, transitive and asymmetric binary relation. Once an alternative is unfeasible in a certain set, then this will remain unchanged in any superset of that set. The authors show that the limited consideration model is fully characterized by a single axiom called Limited Consideration WARP (LC-WARP). This property implies WWARP such that every CLC is also a CTC which suggests that a transformation of the CLC to choice correspondences is presumably also a generalized CTC.

2.6 Bibliography

Apesteguia, J. and Ballester, M. A. (2005). Minimal books of rationales. ftp://ftp.econ.unavarra.es/pub/DocumentosTrab/DT0501.PDF.

Au, P. H. and Kawai, K. (2011). Sequentially rationalizable choice with transitive rationales. *Games and Economic Behavior*, 73(2):608–614.

Cherepanov, V., Feddersen, T., and Sandroni, A. (2013). Rationalization. *Theoretical Economics*, 8(3):775–800.

Danan, E. (2003). Revealed cognitive preference theory. *Technical report, EUREQua, Université de Paris I.*

Eliaz, K. and Ok, E. A. (2006). Indifference or indecisiveness? choice-theoretic foundations of incomplete preferences. *Games and economic behavior*, 56(1):61–86.

Hill, B. (2012). Confidence in preferences. *Social Choice and Welfare*, 39(2):273–302.

Kalai, G., Rubinstein, A., and Spiegler, R. (2002). Rationalizing choice functions by multiple rationales. *Econometrica*, 70(6):2481–2488.

Kreps, D. M. (1988). *Notes on the theory of choice.* Boulder: Westview Press.

Lleras, J. S., Masatlioglu, Y., Nakajima, D., and Ozbay, E. Y. (2014). When more is less: Limited consideration. Working paper, Michigan University.

Mandler, M. (2009). Indifference and incompleteness distinguished by rational trade. *Games and Economic Behavior*, 67(1):300–314.

Manzini, P. and Mariotti, M. (2007). Sequentially rationalizable choice. *American Economic Review*, 97:1824 – 1839.

Manzini, P. and Mariotti, M. (2012). Categorize then choose: Boundedly rational choice and welfare. *Journal of the European Economic Association*, 10(5):1141–1165.

Masatlioglu, Y., Nakajima, D., and Ozbay, E. Y. (2012). Revealed attention. *American Economic Review*, 102(5):2183–2205.

Rubinstein, A. and Salant, Y. (2006). A model of choice from lists. *Theoretical Economics*, 1(1):3–17.

Rubinstein, A. and Salant, Y. (2008). (A, f): Choice with frames. *The Review of Economic Studies*, 75(4):1287–1296.

Rubinstein, A. and Salant, Y. (2012). Eliciting welfare preferences from behavioural data sets. *The Review of Economic Studies*, 79(1):375–387.

Savage, Leonard, J. (1954). The foundations of statistics. *NY, John Wiley*, pages 188–190.

Sen, A. (1993). Internal consistency of choice. *Econometrica: Journal of the Econometric Society*, pages 495–521.

Shafir, E., Simonson, I., and Tversky, A. (1993). Reason-based choice. *Cognition*, 49(1):11–36.

Chapter 3

Managing anticipation and reference-dependent choice

This chapter is based on joint work with Christopher Kops.

Abstract. Anticipation of future consumption may provide pleasure and pain in the present. Recent studies from neurobiology support this view. Building on this evidence, our paper develops a model of individual decision-making under risk where the decision maker derives utility from both standard future consumption and non-standard present anticipation of such consumption. When future consumption is risky, anticipation may range from rational expectations to the narrow focus of dreaming or worrying about a single outcome. On the other hand, anticipation also sets a reference point for consumption. We define a new solution concept, characterize it on the level of choice data and identify the subjective parameters of our model.

Keywords: reference-dependent preferences, reference point, consideration sets, anticipatory utility, information aversion, asymmetric matching pennies

JEL Classification Numbers: JEL Classification: D11, D81, D91

3.1 Introduction

Imagine a decision maker (DM) purchasing a ticket to the state lottery. The jackpot is at a record high: \$1.5 billion. Quite likely, the DM savors the possibility of becoming rich to the degree that she may become slightly upset when she looses. Likewise, it is conceivable that a DM who foregoes to invest in home insurance dreads a potential disaster, until the end of hurricane season may bring her more than a relief. Indeed, recent neurobiological studies show that anticipating future outcomes may produce pleasure and pain in the present (e.g., Berns et al., 2006, 2007; Schmitz and Grillon, 2012). In turn, anticipation is known to affect the joy from actual consumption (Abeler et al., 2011; Baillon et al., 2020). We incorporate this evidence into a theory that generalizes Köszegi and Rabin (2006), define new equilibrium concepts, characterizes them on the level of choice data and identify the subjective parameters of our theory.

Earlier studies have suggested different generalizations of expected utility theory. Among others, these generalizations allow to model loss aversion (Kahneman and Tversky, 1979; Shalev, 2000; Köszegi and Rabin, 2006; Heidhues and Kőszegi, 2008; Kőszegi, 2010), regret aversion (Sugden, 1993), or, disappointment aversion (Bell, 1985; Loomes and Sugden, 1986; Gul, 1991). Prominent examples of such theories explicitly rely on the idea that DMs evaluate outcomes as gains and losses compared to some reference point. Sometimes expectations are taken as the reference point (Köszegi and Rabin, 2006), at other times it is the (augmented) status quo (Kahneman and Tversky, 1979). Experimental evidence on what shapes the reference point is mixed. Some evidence is consistent with reference points being based on people's expectations (Abeler et al., 2011), other evidence is clearly not (Baillon et al., 2020), favoring security levels such as the maximum of the minimal outcomes.

What separates our approach from these earlier generalizations is our new solution concept. First, it allows the DM's reference point for some lottery to be any convex combination of the outcomes possible under that lottery. The reference point may, for instance, be equal to the rational expectation or the lottery's minimum (or maximum) outcome. This way, our model helps to reconcile the theory of reference-dependent preferences with the mixed experimental evidence (Abeler et al., 2011; Baillon et al., 2020). Second, when the DM commits to her choice after anticipation, our solution concept imposes the following consistency on choice: When the DM chooses a lottery, no convex combination of its possible outcomes, as the choice of anticipation, should make her want to choose otherwise. Our characterization and identification results show that the resulting model is still falsifiable. Sharing the concerns that "reference-dependent preferences are inherently difficult to test, as expectations are hard to observe in the field" (Abeler et al., 2011, p.470), we establish these results on the level of observable choice data.

Such characterization and identification results establish to what extent meaning can be inferred from choice data and, at the same time, to what extent the parameters of our model are meaningful in terms of behavior (Dekel and Lipman, 2010). While central to decision theory, identification also provides the grounds on which to draw applications and policy implications (Spiegler, 2008). To further substantiate this claim and suggest possible applications, a few examples may be instructive at this point. Their formalizations are delayed to Section 3.2.4.

Example 3.1.1 (Possibility vs. probability) *Harless and Camerer (1994) show that expected utility theory (EUT) often provides a poor fit when individuals choose between gambles with different support. Specifically, EUT explains behavior rather well when outcome probabilities change from* 24% *to* 25%*, or, from* 60% *to* 61%*, but fails to do so when probabilities change from* 0% *to* 1%*, or, from* 99% *to* 100%*. As our model ties the reference point to the possibility of outcomes, it can account for such behavior.*

Example 3.1.2 (Strategic interactions & anticipation) *In strategic interactions, individuals have to consider what actions or strategies their opponents may settle for. Experiments by Goeree and Holt (2001) on the matching pennies game show that play conforms with Nash-equilibrium predictions when the game is symmetric, but deviates from it systematically in asymmetric instalments of the game. Anticipation in the form of savoring a high payoff or dreading a low one, as is possible in our model, may rationalize such behavior.*

Example 3.1.3 (Information aversion) *Huntington's disease (HD) is a single-gene disorder. Individuals with one parent with HD face a* 50% *chance of inheriting the mutated gene and developing the disease. This means that predictive testing provides significant value. But, extensive evidence shows that at-risk individuals decline to undergo it*

(Evers-Kiebooms et al., 2002; Oster et al., 2013a). Data from self-reports (Oster et al., 2013b) identifies overly optimistic beliefs as a potential driver for such behavior. When anticipation produces pleasure and pain in the present holding such beliefs and stretching the anticipation period may become a rational thing to do.

Example 3.1.4 (Defensive pessimism) *People who score high on standardized anxiety tests have been found to deliberately set significantly lower expectations for themselves, in an effort to prevent the possibility of failure and potential threats to their self-esteem. Norem and Cantor (1986) refer to this phenomenon as* defensive pessimism*. The framework that we lay out here shows how such behavior can be optimal for individuals prone to worrying about future outcomes.*

The rest of the paper is organized as follows. Section 3.2 presents our model of anticipation-based and reference-dependent preferences, defines our equilibrium concepts and formalizes the examples above. Section 3.3 presents a comparable model based on choice and shows that it is equivalent to the model presented in Section 3.2. Section 3.4 provides the behavioral foundation, and Section 3.5 the identification of the model parameters. Section 3.6 discusses our findings. Finally, Section 3.7 concludes.

3.2 A model of anticipation-based and reference-dependent preferences

3.2.1 Primitives

Let X be a finite set of simple (objective) lotteries on some finite set of outcomes Y, where we denote typical elements of Y by x, y and z, and let $\Delta := \Delta(Y)$ be the set of all simple (objective) lotteries on Y. Clearly, for $|Y| \geq 2$, it holds that $X \subset \Delta$, since X is finite. We interpret elements of Δ denoted by r as specifying the lotteries over anticipatory outcomes, and elements of X denoted by p, q and w as lotteries over physical outcomes. Further, for any lottery $p \in X$, $\text{supp}(p)$ denotes the support of p, i.e., the set of outcomes that are possible under the (potentially degenerate) lottery p. If p is the degenerate lottery assigning a probability of 1 to the outcome x and 0 to all other outcomes in X, then

$\text{supp}(p) = \{x\}$. $\mathcal{P}(X)$ denotes the set of nonempty subsets of X with typical elements R, S and T. We refer to these as choice problems.

3.2.2 The model

Since anticipation plays such a key role in our model, the timing of decision-making and consumption becomes a crucial aspect of our model. Formally, we consider a two-period decision problem. In the initial phase of period 1, the DM chooses two lotteries. One is a lottery over physical outcomes to be realized in period 2 and the other is a lottery over anticipatory outcomes to be consumed in period 1.

To address the inherent commitment issue, the next section discusses the following two interpretations of the aforementioned setup: (i) the DM commits to her choice at the time of making the decision and (ii) the DM commits to her choice "shortly before" the lottery over physical outcomes is realized, and hence after deriving anticipatory utility. The equilibrium concepts that we then derive for these two interpretations spell out the requirements for the tuple (r, p) of anticipatory lottery r and physical lottery p to be a feasible choice. A requirement that both concepts share imposes the following rationality constraint on the lottery r: In any equilibrium, the DM can only anticipate what under her choice of a lottery over physical outcomes is theoretically possible. Formally, we can express this as $\text{supp}(r) \subseteq \text{supp}(p)$. For instance, this requirement rules out that the DM dreams about winning the state lottery without purchasing a ticket for it. Furthermore, the DM's choice of r also serves as a reference point for her period 2 consumption of a physical outcome and the utility she derives from it.

Formally, we consider a DM who has preferences $\succsim$ on $\Delta \times X$. We assume that her anticipation-preferences can be represented by a period 1 von Neumann-Morgenstern utility function defined on Δ and her consumption-preferences, likewise, by a period 2 von Neumann-Morgenstern-type utility function defined on $\Delta \times X$. Specifically, given the DM's period-2-choice of a lottery over physical outcomes is $p \in X$ and her period-1-choice of a lottery over anticipatory outcomes is $r \in \Delta$, we assume that her utility is given by

$$\begin{aligned} U(r,p) &= \zeta U_1(r) + \delta U_2(r,p) \\ &= \zeta \sum_{y \in \text{supp}(r)} r(y) u_1(y) + \delta \sum_{z \in \text{supp}(p)} p(z) \sum_{y \in \text{supp}(r)} r(y) u_2(y,z) \end{aligned} \tag{3.1}$$

where u_1, u_2 are continuous Bernoulli utility functions on Y and $Y \times Y$, respectively, and ζ and δ measures the weight of anticipatory and consumption utility, respectively. This is naturally a very general specification. As our main focus is studying how anticipation affects a DM that that attains gain-loss utility, we formulate a specification that extends that of Köszegi and Rabin (2006) in our domain below.[1]

$$U(r,p) = \zeta \sum_{y \in \text{supp}(r)} r(y)u(y) + \delta \sum_{z \in \text{supp}(p)} p(z) \sum_{y \in \text{supp}(r)} r(y) \cdot (u(z) + \mu(u(z) - u(y))) \quad (3.2)$$

where $u = u_1$ and where μ is a gain-loss-function that is increasing with $\mu(0) = 0$. For our exemplary applications here, we focus on the following linear form that extends the one of Köszegi and Rabin (2006),

$$U(r,p) = \zeta \sum_{y \in \text{supp}(r)} r(y) \cdot y + \delta \sum_{z \in \text{supp}(p)} p(z) \sum_{y \in \text{supp}(r)} r(y) \cdot (z + \mu(z - y)) \quad (3.3)$$

where μ is given by

$$\mu(x) = \begin{cases} \eta x, & \text{if } x > 0 \\ \eta \lambda x, & \text{otherwise,} \end{cases} \quad (3.4)$$

and where η is the weight the DM attaches to gain-loss utility, and $\lambda > 1$ is the DM's "coefficient of loss aversion".

3.2.3 Equilibrium concepts

To elaborate on the commitment issue, consider our state lottery example in more detail. Suppose the DM sits at home and ponders the question of whether to go to the local kiosk and buy a ticket for the state lottery. Further, suppose she does all of this shortly before the lottery is resolved. In this case, the DM will derive anticipatory utility already at home if she decides to go and buy the ticket. Note, however, that once at the kiosk she may still renege on her decision to buy the lottery ticket. To address this, our definition of a managing anticipations equilibrium (MAE) directly imposes a feasibility restriction on what the DM can choose when she can only commit to her choice after anticipation.

Definition 3.1 (MAE) *Given a two-period choice problem $S \in \mathcal{P}(X)$, the tuple $(r_p^*, p) \in \Delta \times S$ is a managing anticipations equilibrium (MAE) if*

[1] All of our results, however, also hold for the general specification in Equation (3.1).

1. $\text{supp}(r_p^*) \subseteq \text{supp}(p)$*,*

2. $U(r_p^*, p) \geq U(r_p, q)$*, for all* $(r_p, q) \in \Delta \times S$ *with* $\text{supp}(r_p) \subseteq \text{supp}(p)$

In other words, a tuple of lotteries, (r_p^*, p), is a MAE for some two period choice problem S if it satisfies the following three conditions. First, the consumption lottery p must be available as a choice, i.e., $p \in S$. Second, if the DM chooses p from choice problem S, then the only outcomes she may anticipate are those that are possible under her choice of p, i.e., $\text{supp}(r_p^*) \subseteq \text{supp}(p)$. Third, given the DM chooses p, there should not be anything that she can anticipate from this choice that changes her mind as to whether she wants to choose p.

At first glance, the MAE concept may seem to allow for almost all types of behavior one can think of. This is far from true. Consider, for instance, a DM who is asked to decide whether or not she wants to purchase, at a cost of $b > 0$, a lottery that pays out $a > 0$ with probability p and nothing otherwise. If utility u_1 is concave and $u_2 = u_1 + \mu$, with μ as defined in Equation (3.4), then it is straightforward to verify that any MAE involves not buying the lottery if the lottery's expected value is negative. Indeed, in this case, purchasing the lottery and anticipating to win (resp., lose) is dominated by not buying the lottery and anticipating to win (resp., lose). The second part of Definition 3.1, then rules out that purchasing the lottery is a MAE. This illustrates how in our example above the DM potentially will not buy the lottery if she commits to her decision only shortly before the lottery is resolved.

Our definition above does not guarantee that the MAE unique. In fact, multiple MAEs with different utility levels may arise. This naturally suggests a refinement of the MAE concept. This refinement imposes that the DM will not just choose any or all of the MAEs, but her most preferred ones, leading to the equilibrium concept that we name preferred managing anticipation equilibrium (PMAE). We define it in the following way.

Definition 3.2 (PMAE) *Given a two-period choice problem* $S \in \mathcal{P}(X)$*, the tuple* $(r_p^*, p) \in \Delta \times S$ *is a preferred managing anticipations equilibrium (PMAE) if*

1. (r_p^*, p) *is a MAE, and*

2. $U(r_p^*, p) \geq U(r_q^*, q)$*, for all MAEs* (r_q^*, q) *with* $(r_q^*, q) \in \Delta \times S$

In other words, the tuple (r_p^*, p) is a PMAE for some two-period choice problem S if it is a MAE and weakly preferred to any other possible MAE.

To motivate our final equilibrium concept, we return to the commitment issue from before. In a MAE, the DM anticipates the decision situation, but commits to her actual choice shortly before the lottery over consumption outcomes realizes, so that the reference point stays fixed. In the equilibrium concept we introduce next, the choice-acclimating managing anticipations equilibrium (CMAE), the DM commits to her choice "far in advance" so that the reference point may adapt. This is analogous to the difference between the personal equilibrium (PE) and choice-acclimating PE (CPE) in Kőszegi and Rabin (2007).

Definition 3.3 (CMAE) *Given a two-period choice problem $S \in \mathcal{P}(X)$, the tuple $(r_p^*, p) \in \Delta \times S$ is a choice-acclimating managing anticipations equilibrium (CMAE) if*

1. $\text{supp}(r_p^*) \subseteq \text{supp}(p)$,

2. $U(r_p^*, p) \geq U(r_q, q)$, *for all* $(r_q, q) \in \Delta \times S$ *with* $\text{supp}(r_q) \subseteq \text{supp}(q)$

In other words, the tuple (r_p^*, p) is a CMAE for some two-period choice problem S if it satisfies the first two conditions of a MAE and if the utility from (r_p^*, p) is higher than that from any other tuple (r_q, q) that also satisfies the two first conditions of a MAE.

Note that the definition of the CMAE is less restrictive than that of the MAE, in the sense that the DM can choose according to her preference over $\Delta \times X$, as long as the reference lottery is in the support of the physical lottery. This is so, because the DM can commit to her choice at the time of choosing. Indeed, in our state lottery example, the DM only needs to compare the choice of buying the ticket and dreaming about winning to not buying and anticipating just that. This scenario is comparable to a situation where the DM finds herself in the kiosk and decides whether to buy the ticket for tomorrow's state lottery without having anticipated this purchase in advance. Thus, this concept hints at an "impulse" buying effect, in the sense that more alternatives are feasible when the choice problem has not been anticipated in advance.

3.2.4 Examples

We now present a couple of examples which show how the equilibrium concepts can be applied and that elaborate on the examples presented in the Introduction. The first example considers the famous Allais paradox. Typical Allais-type choices are inconsistent with expected utility theory. Probability weighting as proposed in prospect theory (Kahneman and Tversky, 1979) is usually invoked to explain these choices. Our approach in this paper provides an alternative explanation for Allais-type behavior that is based on anticipation being an important source of utility and the DM being considerably loss-averse. Such a DM is prone to worrying a lot about a lottery resulting in a bad outcome, no matter how unlikely this is to happen (as long as it is still possible, of course). In other words, such a DM wants to safeguard herself from feeling any disappointment in the future.

Example 3.2.1 (Possibility vs. probability (cont'd)) *Consider a DM who is confronted with the two choice problems of the Allais Paradox. First, she is asked to choose between* $p = [\$1M]$ *and* $q = (\$1M, .89; \$0, .01; \$5M, .1)$*, and, second, between* $\tilde{p} = (\$0, .89; \$1M, .11)$ *and* $\tilde{q} = (\$0, .9; \$5M, .1)$*. Typical Allais choices are* p *from* $\{p, q\}$ *and* $\tilde{q}$ *from* $\{\tilde{p}, \tilde{q}\}$*. If anticipation is a rather important source of pleasure and pain for the DM, say* $\zeta > \delta$*, and she is considerably loss averse, i.e.,* λ *is large enough, then it is straightforward to show that the typical Allais choices can be the unique CMAE and PMAE. The mere possibility of* $\$0$ *and* $\$5M$ *drives here choices of* p *from* $\{p, q\}$ *and* $\tilde{q}$ *from* $\{\tilde{p}, \tilde{q}\}$*, respectively.*

The next example illustrates that a model of anticipation is applicable in situations incorporating strategic considerations. In a series of experiments by Goeree and Holt (2001), participants were asked to interact in several strategic situations. Two of these situations took the form of the Matching Pennies Game depicted in Figure 3.1.

Figure 3.1: Symmetric (left) and asymmetric (right) versions of a Matching Pennies Game between Player 1 (P_1) and Player 2 (P_2)

		P_2	
		L (48%)	R (52%)
P_1	T (48%)	80,40	40,80
	B (52%)	40,80	80,40

		P_2	
		L (16%)	R (84%)
P_1	T (96%)	320,40	40,80
	B (4%)	40,80	80,40

Note: Percentages in brackets indicate how many participants in the experiment by Goeree and Holt (2001) chose which pure strategy.

With 48% and 52%, average play in the symmetric Matching Pennies Game was very close to the mixed-strategy Nash equilibrium of 50–50 as predicted by standard game theory. In the asymmetric Matching Pennies Game, however, average play of Player 1 with 96% for T and 4% for B was far off the equilibrium play. The following example illustrates that PMAE cannot only account for these choices, it also provides an intuitive explanation for why the observed play may be an equilibrium after all. This explanation rests again on loss-aversion and anticipation being a non-negligible source of utility.

Example 3.2.2 (Strategic interactions & anticipation (cont'd)) *There are two players, Player 1 and Player 2, with identical preferences, represented by Equation (3.3). Their choice behavior is governed by PMAE*[2]*. Further, let them put almost equal weight on the anticipatory and the consumption part of their utility function, i.e.,* $\delta = .52 > .48 = \zeta$*, a weight of* $\eta = .908$ *on gain-loss utility and let their coefficients of loss aversion be equal to* $\lambda = 1.12$*. Then, in the asymmetric Matching Pennies Game in Figure 3.1, Player 1 playing T with probability* $.96$ *and Player 2 playing L with probability* $.16$ *is a mixed-strategy equilibrium built on PMAEs. This is so, because, while Player 1 dreams about receiving 320, when playing T, he worries about receiving 40, when playing B. Taking this into account, Player 2 worries about receiving 40, when playing L, and dreams about receiving 80, when playing R.*

The subsequent example we present shows how anticipatory utility is capable of rationalizing situations of information aversion.

Example 3.2.3 (Information aversion (cont'd)) *Consider an at-risk individual with one parent suffering from HD who is considering whether to get tested for the mutated gene. Not getting tested is equivalent to choosing the lottery* $p = (a, .5; b, .5)$ *where* $a > 0$ *is the outcome of no HD and* $b < 0$ *the outcome of having HD. Choosing to get tested will resolve all uncertainty, and will thus make it impossible to anticipate anything else other than the true state. However, when facing the choice problem, the state is not known. Let the degenerate lottery* $q = [.5a + .5b]$ *be the outcome of getting tested. With linear utility,*

[2] We focus on the case of PMAE here. However, as one may argue that the CMAE is more appropriate in the experimental setup, we note that it is straightforward to show that similar parametric definitions lead to the same conclusion.

anticipating not having HD and not getting tested is a MAE if both

$$\zeta a + \delta(.5 * a + .5 * (b + \eta\lambda(b - a))) \geq \zeta a + \delta(.5 * a + .5 * b + \eta\lambda(.5 * b - .5 * a))$$

and

$$\zeta a + \delta(.5 * a + .5 * (b + \eta\lambda(b - a))) \geq \zeta b + \delta(.5 * a + .5 * b + \eta(.5 * a - .5 * b))$$

Furthermore, this option is guaranteed to be an PMAE if, in addition, anticipating not having HD not getting tested is better than getting tested. This holds if

$$\zeta a + \delta(.5 * a + .5 * (b + \eta\lambda(b - a))) \geq \zeta(.5 * a + .5 * b) + \delta(.5 * a + .5 * b)$$

It is straightforward to check that all inequalities are satisfied for $\zeta \geq \delta\eta\lambda$. That is, if the weight on anticipatory utility is large enough. This shows that not getting tested and anticipating not having the disease is an equilibrium. Anticipation driving information-averse choice in this way matches the data on self-reports from at-risk individuals (Oster et al., 2013b) suggesting overly optimistic beliefs as the driver for this information-averse decision.

The final simple example shows how consciously anticipating the worst outcome may be optimal in our theory.

Example 3.2.4 (Defensive pessimism (cont'd)) *Consider an anxious young economist who has presented her research paper at the most pertinent conferences, made sure that her paper is as polished as it can be and has submitted it to a top ranked scientific journal in her field. The publication process she faces is like a lottery that puts low probability weight on her paper being published and a high probability on it being rejected. Let $a > 0$ be the outcome of publication and $b < 0$ the outcome of rejection. Furthermore, let $p \ll .5$ be the probability of publication. In then follows immediately that the young economist will anticipate rejection if she is sufficiently loss averse.*[3]

3.2.5 Comments on equilibrium existence and equivalence

Whereas it follows directly that the set of alternatives that are CMAE in a given choice problem S is always nonempty, it is straightforward to see from Definition 3.1 that situa-

[3] Notice that since the economist has already chosen to submit her paper to the journal, PMAE and CMAE coincide.

tions with no MAE may arise. To address this, we note that a reformulated version of the consistency requirement *limited cycle inequalities* (Freeman, 2017) imposed on the utility function, U, is sufficient for equilibrium existence. Specifically, the following condition suffices:

Definition 3.4 (Limited cycle inequalities) *The function* $U : \Delta \times X \to \mathbb{R}$ *satisfies* limited cycle inequalities *if for any* $p_0, \ldots, p_n \in X, U(r^*_{p_{i-1}}, p_i) > U(r_{p_{i-1}}, p_{i-1})$ *for* $i = 1, \ldots, n$, *then* $U(r^*_{p_n}, p_n) > U(r_{p_n}, p_0)$, *for some* $r^*_{p_{i-1}} \in \text{supp}(p_{i-1})$, *for all* $r_{p_{i-1}} \in \text{supp}(p_{i-1})$, *for some* $r^*_{p_n} \in \text{supp}(p_n)$, *and for all* $r_{p_n} \in \text{supp}(p_n)$.

In words, limited cycle inequalities state that, if there is a sequence of consumption lotteries such that each succeeding lottery in the sequence makes the preceding lottery non-feasible according to the MAE concept, then the first lottery in the sequence cannot block the last lottery. Our results in later sections depend on U satisfying this property.[4]

Next, from the definitions of the equilibrium concepts, the following corollary shows under which conditions (i) the set of MAEs is equivalent to the set of PMAEs and (ii) the set of PMAEs is equivalent to the set of CMAEs.

Corollary 3.1 1. *MAE is equivalent to PMAE if and only if for all* $p, q \in X$, $U(r^*_p, p) \geq U(r_p, q)$ *and* $U(r^*_q, q) \geq U(r_q, p)$ *imply that* $U(r^*_p, p) = U(r^*_q, q)$ *for all* $r_p \in \Delta$ *with* $\text{supp}(r_p) \subseteq \text{supp}(p)$, *for all* $r_q \in \Delta$ *with* $\text{supp}(r_q) \subseteq \text{supp}(q)$, *for some* $r^*_p \in \Delta$ *with* $\text{supp}(r^*_p) \subseteq \text{supp}(p)$, *and for some* $r^*_q \in \Delta$ *with* $\text{supp}(r^*_q) \subseteq \text{supp}(q)$.

2. *PMAE is equivalent to CMAE if and only if for all* $p, q \in X$, $U(r^*_p, p) \geq U(r_q, q)$ *implies that* $U(r^*_p, p) \geq v(r_p, q)$ *for some* $r^*_p \in \Delta$ *with* $\text{supp}(r^*_p) \subseteq \text{supp}(p)$, *for all* $r_p \in \Delta$ *with* $\text{supp}(r_p) \subseteq \text{supp}(p)$, *and for all* $r_q \in \Delta$ *with* $\text{supp}(r_q) \subseteq \text{supp}(q)$.

The corollary firstly states that all MAEs are PMAEs if and only if U is defined such that the utility of any two MAE must be the same. Secondly, it states that all PMAEs are CMAEs if and only if any CMAE is also an MAE.

[4] Note that Freeman (2016) shows that the property is satisfied for functional forms that are typically used for modelling reference-dependent choice.

3.3 Reference-dependent choice

The purpose of this and the following sections is to consider our model of anticipation-based and reference-dependent preferences presented in Section 3.2 on the domain of observables. The idea is that we may only observe the set of available alternatives S and the DM's chosen alternatives. In this domain, we show that the PMAE concept (Definition 3.2) is equivalent to a two-stage choice procedure. In the first stage, a subset of the available alternatives is chosen for consideration based on a filtering process that satisfies well-known internal consistency conditions. Finally, in the second stage, the DM applies her preference relation to select the most preferred of the considered alternatives. The equivalence is further underlined by the fact that the first-stage filtering is equivalent to the MAE concept (Definition 3.1). In addition, it follows immediately that the CMAE concept (Definition 3.3) is equivalent to a DM choosing according to her preference relation without being restricted by a filtering in the first stage.

3.3.1 Primitives

Let X be the set of lotteries over physical outcomes with typical elements denoted by p, q and w as defined in Section 3.2. $\mathcal{P}(X)$ denotes the set of nonempty subsets of X with typical elements R, S and T. A class is a collection of sets. A cover of a set S is a collection of sets whose union contains S as a subset. Formally, if $\mathcal{C} = \{T_i : 1 \leq i \leq n\}$ is an indexed class of sets T_i, then $\mathcal{C}$ is a cover of S if $S \subseteq \bigcup_{i=1}^{n} T_i$. A binary relation R on X is a subset of $X \times X$, where we abbreviate $(p, q) \in R$ by pRq. R is a weak order if it is complete (i.e., pRq or qRp for all $p, q \in X$) and transitive (i.e., pRq and qRw implies pRw for all $p, q, w \in X$). The set of maximal alternatives according to a binary relation R on some set S is given by

$$\max(S; R) = \{p \in S : pRq \text{ if } qRp, \forall q \in S\}$$

We denote the asymmetric component of the binary relation R by P. That is, for any $p, q \in X$, pPq if and only if pRq and $\neg[qRp]$. A correspondence $C : \mathcal{P}(X) \to \mathcal{P}(X)$ is a mapping that for any choice problem $S \in \mathcal{P}(X)$ picks a nonempty subset $C(S) \subseteq S$. It follows that a choice correspondence is induced by a binary relation R, if and only if $C(S) = \max(S; R)$ for all $S \in \mathcal{P}(X)$. In that case, we use the notation $C(S) = C(S; R)$.

3.3.2 The choice procedure

We consider a DM with preferences over X that are captured by a weak order $\succsim$. The starting point of our choice theoretical analysis is a choice correspondence on X. The choice theory that we set up is such that the DM picks the preferred alternatives from all alternatives which she considers for her choice owing to a filtering process. These may not be the best of the available alternatives according to the DM's preferences. Rather, it will be the most preferred alternatives that receives her consideration, i.e., from a subset of the available alternatives. Our goal is to later elicit the DM's preference relation along with her consideration set from choice data alone. This is impossible without any knowledge about how she forms this consideration set. To see this, note that otherwise it would always be possible to claim that the DM only considers the alternatives she chooses and nothing else and that these are equally preferred, such that outside observers can hardly infer anything about the DM's preferences and consideration set. The next definition states our assumptions on referential consideration formally.

Definition 3.5 *A consideration set mapping $\Gamma : \mathcal{P}(X) \to \mathcal{P}(X)$ is a reference-dependence (RD) filter if for any $S \in \mathcal{P}(X)$ and any $p \in S$, it holds that $p \in \Gamma(S)$, whenever there exists a cover $\mathcal{C}$ of S such that $p \in \Gamma(T)$, for all $T \in \mathcal{C}$.*

The idea here draws on that of revealed preference extended to revealed referential consideration. If the DM's consideration and, ultimately, her choice is influenced by reference points, then this influence should be internally consistent on the level of consideration. Given this definition of the filter, we can now introduce the choice procedure that we are proposing in this paper.

Definition 3.6 *A choice correspondence C on X is a reference-dependent choice (RDC) if there exists a weak preference relation $\succsim$ on X and an RD filter Γ such that for any choice problem $S \in \mathcal{P}(X)$, $C(S)$ is the set of $\succsim$-best elements in $\Gamma(S)$, formally*

$$C(S) = C(\Gamma(S); \succsim), \quad \forall S \in \mathcal{P}(X)$$

The definition of the RDC states that, in any choice problem, the DM first uses her reference dependence (RD) filter to limit the set of alternatives. From this limited set, she then applies her preference relation to select the most preferred alternatives. In the next section, we will show the equivalence of (i) the RDC and PMAE and (ii) the RD filter and MAE.

3.3.3 Behavioral equivalence of PMAE and RDC

This section provides a behavioral equivalence result of the PMAE concept and the RDC procedure. Such results enable any outside observer to verify whether choice data coming out of a series of choice problems is consistent with a DM choosing based on anticipation and reference-dependent preferences. To do so, we adopt the predominant view in the literature on reference points (Freeman, 2017) and anticipations (Abeler et al., 2011) in that these elements of our theory are hard if not impossible to observe. Therefore, the primitives on which we provide our behavioral characterizations of the MAE and the PMAE concept is choice over physical outcome-lotteries alone, that is, choice from X. To start things, let Δ be the set of lotteries over anticipatory outcomes, with typical elements denoted by r as defined in Section 3.2. Let $U : \Delta \times X \to \mathbb{R}$ be a reference-dependent utility function as defined in Equation (3.1) satisfying the limited cycle inequalities property as in Definition 3.4. For any alternative $p \in X$, define R_p as follows:

$$R_p = \{r \in \Delta : \mathrm{supp}(r) \subseteq \mathrm{supp}(p)\}$$

that is, R_p is the set of anticipatory choices that are in the support of the alternative p.

Following Definition 3.1, let $C(\cdot; U_{\mathrm{MAE}})$ be a choice correspondence C induced by U in accordance with the MAE concept. That is, for any $S \in \mathcal{P}(X)$, the correspondence is given by

$$C(S; U_{\mathrm{MAE}}) = \{p \in S : U(r_p^*, p) \geq U(r_p, q), \forall q \in S, \forall r_p \in R_p \text{ and some } r_p^* \in R_p\}$$

The following result establishes that the MAE concept is equivalent to choosing directly by the RD filter. As such, the RD filter can be thought of as a consideration set mapping that spotlights the set of feasible alternatives, the ones that are MAE.

Proposition 3.1 *Let X be a set of alternatives and $C : \mathcal{P}(X) \to \mathcal{P}(X)$ a choice correspondence. There exists an RD filter Γ such that $C(S) = \Gamma(S)$ for all $S \in \mathcal{P}(X)$ if and only if there exists a $U : \Delta \times X \to \mathbb{R}$ such that $C(S) = C(S; U_{\mathrm{MAE}})$.*

Proof: Please refer to Section 3.9.1.

The PMAE concept is a natural refinement on the set of MAEs, based on the DM picking the optimal MAE according to her utility, U, conditionally on satisfying the individual rationality requirement. Accordingly, $C(S; U_{\text{PMAE}})$ is defined as follows

$$C(S; U_{\text{PMAE}}) = \{p \in S : U(r_p^*, p) \geq U(r_q, q), \forall q \in C(S; U_{\text{MAE}}), \forall r_q \in R_q \text{ and some } r_p^* \in R_p\}$$

The following proposition establishes the equivalence of RDC and PMAE.

Proposition 3.2 *Let X be a set of alternatives and $C : \mathcal{P}(X) \to \mathcal{P}(X)$ a choice correspondence. There exists a weak preference $\succsim$ and an RD filter Γ such that $C(S) = C(\Gamma(S); \succsim)$ for all $S \in \mathcal{P}(X)$ if and only if there exists a $U : \Delta \times X \to \mathbb{R}$ such that $C(S) = C(S; U_{\text{PMAE}})$.*

Proof: Please refer to Section 3.9.2.

Propositions 3.1 and 3.2 jointly show that the equivalence of the RDC and our model of anticipation-based reference-dependent preferences defined in Section 3.2 is such that the RD filter fully captures the set of alternatives from a choice problem S that are MAE. Then the second-stage application of the weak order in the RDC fully captures the the DM's preference among the alternatives that are MAE. Finally, the simpler CMAE concept merely requires that the DM chooses the best alternative in S according to U. Thus, the definition of $C(S; U_{\text{CMAE}})$ is given by

$$C(S; U_{\text{CMAE}}) = \{p \in S : U(r_p^*, p) \geq U(r_q, q), \forall q \in S, \forall r_q \in R_q \text{ and some } r_p^* \in R_p\}$$

The following straightforward result establishes that the CMAE concept is equivalent to choosing by the weak order, that is, it is equivalent to a special case of the RDC in which all alternatives receives consideration.

Corollary 3.2 *Let X be a set of alternatives and $C : \mathcal{P}(X) \to \mathcal{P}(X)$ a choice correspondence. There exists a weak preference $\succsim$ such that $C(S) = C(S; \succsim)$ for all $S \in \mathcal{P}(X)$ if and only if there exists a $U : \Delta \times X \to \mathbb{R}$ such that $C(S) = C(S; U_{\text{CMAE}})$.*

3.4 Behavioral foundation

Based on our results in Section 3.3, we are now able to provide behavioral characterizations of the RDC and its components. By extension, this provides a characterization of our model of anticipation-based reference-dependent preferences. Such characterizations deem our model as well as its components falsifiable on observable choice data alone. We start our investigation by defining the two axioms which jointly characterizes a DM choosing solely by an RD filter.

Axiom 3.1 (Contraction (Property α))

$$[p \in C(T) \ \textit{and} \ S \subset T, \ p \in S] \Rightarrow [p \in C(S)]$$

Contraction simply requires that if an alternative p is chosen from some set T, then it must be chosen in any subset S that contains p. Thus, contraction imposes the behavioral restriction that removing irrelevant alternatives (or relevant, that is, alternatives that are also chosen from T) do not alter the choice of p.

Axiom 3.2 (Expansion (Property γ^*)) *For all $S, T \in \mathcal{P}(X)$,*

$$[p \in C(S) \cap C(T)] \ \Rightarrow \ [p \in C(S \cup T)]$$

Expansion says that if an alternative is chosen from two sets, S and T, then this alternative will also be chosen from the union of these sets, i.e., from $S \cup T$. It is a well-known axiom that plays a vital role in the characterization of the rational shortlist method (RSM) of Manzini and Mariotti (2007).[5]

The next proposition shows that a DM choosing based on a consideration set mapping being an RD filter is equivalent to satisfying contraction consistency and expansion consistency.[6]

Proposition 3.3 *Let X be a finite set of alternatives and $C : \mathcal{P}(X) \to \mathcal{P}(X)$ a choice correspondence. There exists an RD filter Γ such that $C(S) = \Gamma(S)$ for all $S \in \mathcal{P}(X)$ if and only if C satisfies Contraction and Expansion.*

Proof: Please refer to Section 3.9.3.

We now provide a behavioral characterization of the RDC procedure. By extension, this serves as a characterization of the PMAE. Let us reiterate here that such a characterization enables any outside observer to verify whether choice data is consistent with

[5] Note that the expansion axiom in Manzini and Mariotti (2007) is defined for choice function. The expansion axiom defined here is a natural extension for choice correspondences (Sen, 1971).

[6] Note, that it is a well-known fact that contraction and expansion together are logically equivalent to the weak axiom of revealed preferences (WARP) (Sen, 1971; Tyson, 2008).

the RDC procedure or not, without knowing the DM's RD filter and preferences. Before we do this, note that because choices supported by CMAE are behaviorally equivalent to choosing according to a weak order, the observed behavior is rational in the neoclassical sense.[7] To characterize the procedure, we will need to introduce two additional axioms.

Axiom 3.3 (Weak WARP (WWARP)) *For all $S, T \in \mathcal{P}(X)$,*

$$[\{p, q\} \subset S \subseteq T, q \notin C(\{p, q\}) \text{ and } p \in C(T)] \;\Rightarrow\; [q \notin C(S)]$$

This axiom is a weakening of the well-known weak axiom of revealed preferences (WARP) and is an essential part of the behavioral characterizations of several choice procedures such as the RSM (Manzini and Mariotti, 2007), categorize then choose (CTC) (Manzini and Mariotti, 2012), and the rationalization model (Cherepanov et al., 2013).[8] WWARP says that if an alternative p is chosen from a set T containing q and uniquely chosen when q is the only other available alternative, then q cannot be chosen from any subset S of T that contains p.

Axiom 3.4 (No Reversible Binary Cycles (NRBC)) *For all $p_1, \dots, p_{n+1} \in X$ and $S_1, \dots, S_{n+1} \in \mathcal{P}(X)$ with $p_i \in S_i$, for all $i = 1, \dots, n+1$*

$$[p_i \in C(\{p_i, p_{i+1}\}),\, p_{i+1} \in C(S_i), \text{ and } p_1 \in C(S_{n+1}), \forall i = 1, \dots, n] \;\Rightarrow\; [p_1 \in C(\{p_1, p_{n+1}\})]$$

NRBC requires that there are no pairwise cycles of choice which are reversed in larger sets. It is more restrictive than the condition of No Binary Cycles (Manzini and Mariotti, 2007) that requires that there are no pairwise cycles of choice whatsoever.

The following result then establishes that Expansion, WWARP, and NRBC form the behavioral characterization of RDC.

Theorem 3.1 *Let X be a finite set of alternatives and $C : \mathcal{P}(X) \to \mathcal{P}(X)$ a choice correspondence. There exists a weak preference $\succsim$ and an RD filter Γ such that $C(S) = C(\Gamma(S); \succsim)$ for all $S \in \mathcal{P}(X)$ if and only if C satisfies Expansion, WWARP, and NRBC.*

Proof: Please refer to Section 3.9.4.

[7] That is, there exists a function $U : \Delta \times X \to \mathbb{R}$ such that choices are induced by U in accordance with CMAE if and only if they satisfy WARP.

[8] Note again that our WWARP axiom is a natural extension for choice correspondences introduced in Armouti-Hansen and Kops (2018).

3.5 Identification

The RDC is based on two key parameters that enter into the DM's decision-making procedure: her preferences and her consideration set. In the last section, we identified three testable conditions that can be applied to any given choice data to determine whether this data can be thought of as resulting from an RDC procedure. Now, suppose we have choice data that is consistent with the RDC logic. The question that we address in this section is about the extent to which the two key parameters of the RDC procedure can be uniquely identified from such data.

We first consider the question of identification of the DM's preferences. In contrast to rational choice theory, with theories of bounded rationality like the RDC, there may be multiple possible preferences which can rationalize the same choice data. To check whether the DM ranks p above q, it, thus, seems natural to check whether every possible representation of choices ranks p above q (Masatlioglu et al., 2012). The following definition is useful to organize the discussion.

Definition 3.7 *Let C be an RDC. We say that p is revealed to be "weakly preferred" to q by the DM, if for any $(\succsim, \Gamma)$ that is part of a RDC representation of C, we have $p \succsim q$. Furthermore, we say that p is revealed to be "strictly preferred" to q by the DM if for any $(\succsim, \Gamma)$ that is part of a RDC representation of C, we have $p \succsim q$ and $\neg[q \succsim p]$.*

Checking every possible representation of choices may not be a very practicable method. Fortunately, there is a simpler way to identify the DM's preferences. To capture this idea, we define the following binary relation R on X via

$$pRq \text{ if } p \in C(\{p,q\}) \text{ and } q \in C(S), \text{ for some } S \in \mathcal{P}(X) \text{ with } \{p,q\} \subseteq S$$

The relation R can be interpreted as a revealed preference relation that can be *directly* elicited from choice data. This is so because q is chosen from S and, thus, considered. Further, we know from the definition of the RD filter that $q \in C(S)$ and $\{p,q\} \subseteq S$ implies that q receives consideration in $\{p,q\}$, as well. Hence, $q \in \Gamma(\{p,q\})$ and thus, by definition of RDC, $p \in C(\{p,q\})$ implies that $p \succsim q$.

Next, define R^* to be the transitive closure of R. It also follows that if pR^*q, then p is revealed to be weakly preferred to q. Loosely speaking, this is true because if pRw and

wRq (and hence pR^*q) for some w then, since the underlying weak preference relation defining a RDC representation is transitive, it follows that p is revealed to be weakly preferred to q even when pRq is not directly revealed from choices. The question remains whether R^* really captures all revealed preferences and, at the same time, not more than that. The next proposition establishes that R^* really is the revealed preference.

Proposition 3.4 *Let C be an RDC. Then p is revealed to be weakly preferred to q if and only if pR^*q.*

Proof: Please refer to Section 3.9.5.

We now address the issue of identifying whether an alternative p is revealed to be strictly preferred to q. Let P^* to be the transitive closure of the binary relation P, which we defined as follows

$$pPq \text{ if } C(\{p,q\}) = \{p\} \text{ and } q \in C(S), \text{ for some } S \in \mathcal{P}(X) \text{ with } \{p,q\} \subset S$$

The intuition is as follows. Since q is chosen from S, it is considered in S. Since q is considered in S, it is considered in all its subsets as well, i.e., also in $\{p,q\}$. Since p is chosen from $\{p,q\}$ but q is not, we must have $p \succsim q$ and $\neg[q \succsim p]$ by definition of the RDC. The following corollary follows immediately.

Corollary 3.3 *Let C be an RDC. Then p is revealed to be strictly preferred to q if and only if pP^*q.*

Under the theory of rational choice, more can never be less in terms of individual welfare. That is, if $S \subset T$, no rational DM will strictly prefer any alternative in $C(S)$ to any alternative in $C(T)$. Under theories of bounded rationality, more can indeed be less. Based on this the next corollary boils the observation that more may be less down to an easily testable statement.

Corollary 3.4 *More is less (in terms of welfare) if $C(S \cup C(T)) \neq C(T)$, for some $S \subset T$.*

In other words, less alternatives can be more in terms of the DM's well-being. Specifically, anytime it holds that $C(S \cup C(T)) \neq C(T)$, for some $S \subset T$, then the DM is better off choosing from the smaller set S, compared to choosing from the larger set T.

Next, we consider the question of identification of the DM's consideration set. Again, to check whether the DM considers p at S for her choice, it seems natural to check whether every possible RDC representation of the DM's choices specifies that p receives the DM's consideration at S (Masatlioglu et al., 2012). In a similar way as before, the following definition is useful to organize the discussion.

Definition 3.8 *Let C be an RDC. We say that p is revealed to receive consideration at S by the DM, if for any $(\succsim, \Gamma)$ that is part of a RDC representation of C, we have $p \in \Gamma(S)$.*

It turns out, there also exists a simple way to identify the DM's revealed consideration set. To this end, we define the following consideration set Γ^* on $\mathcal{P}(X)$ via

$$p \in \Gamma^*(S) \text{ if } p \in S \text{ and } p \in C(T) \text{ for some } T \in \mathcal{P}(X) \text{ with } S \subseteq T$$

Again, the question remains whether Γ^* really captures all revealed consideration and, at the same time, not more than that. The next proposition establishes that Γ^* really is the revealed consideration set.

Proposition 3.5 *Let C be an RDC. Then p is revealed to receive consideration at S if and only if $p \in \Gamma^*(S)$.*

Proof: Please refer to Section 3.9.6.

Under the theory of rational choice, more implies more in terms of alternatives to choose from. That is, if $S \subset T$, no rational DM will consider (weakly) less alternatives for her choice from T than she does for her choice from S. Again, under theories of bounded rationality, more can indeed be (weakly) less. Specifically, for the RDC, more does not always imply (strictly) more in terms of consideration. The next corollary states this formally.

Corollary 3.5 *More does not imply (strictly) more (in terms of consideration). That is, if $C(S) \neq C(T)$, for some $S \subset T$ with $C(T) \subseteq S$, then $\Gamma(T) \subseteq \Gamma(R)$, for some $R \subsetneq T$.*

In other words, a larger set does not automatically imply that the DM also considers more alternatives for her choice. Specifically, if $C(S) \neq C(T)$, for some $S \subset T$ with $C(T) \subseteq S$, then there exists a subset R of T such that all alternatives that the DM considers for her choice when choosing from T, she also considers for her choice when choosing from R.

3.6 Discussion

As we have shown, the definition of the RDC is behaviorally equivalent to the PMAE concept. We find that this formulation, with a reference-dependent consideration filter, is a natural translation of the model in terms of choice. On the other hand, there naturally exist alternative definitions of the choice procedure that retain its behavioral equivalence to PMAE. As Freeman (2017) shows for the preferred personal equilibrium (PPE), we could instead define the RDC as a special case of an extension of the RSM of Manzini and Mariotti (2007) to the case of choice correspondences. In the RSM, the DM arrives at her choice by an ordered sequential elimination of inferior alternatives based on two rationales P_1 and P_2, which are asymmetric binary relations. As is shown in Armouti-Hansen and Kops (2018), the RSM extended to the case of choice correspondences is characterized by Expansion and WWARP. Since the characterization of PMAE requires NRBC in addition, it follows that the binary relations of the RSM need to be further restricted to achieve equivalence. Thus, we could instead define the RDC as a special case of an extension of the RSM in which P_1 is asymmetric and the second stage rationale, P_2, is complete and transitive.[9] In relation to this, a notable point is that, when only choices are observable, the equilibrium concepts from this anticipation-based reference-dependent model are indistinguishable from the equilibrium concepts of Köszegi and Rabin (2006); Kőszegi and Rabin (2007) on our domain. In particular, the rules that governs admitable choice behavior of the PE and PPE are equivalent to that of the MAE and PMAE, respectively. Unsurprisingly, the same holds true for the CPE and CMAE.

Naturally, our proposed model in Section 3.2 can be extended on several accounts in future research. Firstly, the domain of simple lotteries may be extended to include more complex objects. Secondly, the number of periods may be extended to both highlight and differentiate between anticipation prior to the DM's commitment and anticipation after commitment, but prior to the consumption lottery's resolution. Thirdly, one may extend our model to allow for situations in which a second party may influence the DM's choice of anticipation. In particular, from a marketing perspective, if exposing potential consumers to a certain state influences the DM's anticipation, this may, in turn, increase her willingness to pay. Another potential application would be that of a transformational

[9] This follows from Proposition 2 in Freeman (2017).

leader. If speaking about positive visions affect employees' anticipation, this in turn may increase their effort.

As a final point, we note that the definition of the RDC with its weak order is a natural extension of the case of a strict linear order. We use it because it imposes less restrictions on the utility function in order to establish equivalence between PMAE and RDC. As it is common practice in choice theory to restrict the analysis to the case of a choice function, we briefly address this here. The characterization of the special case where the RDC is a choice function is given by the axioms of Expansion and WWARP defined in Manzini and Mariotti (2007, 2012); Cherepanov et al. (2013) joint with the following reformulation of NRBC: $[c(p_i, p_{i+1}) = p_i,\ c(S_i) = p_{i+1}, \text{and } c(S_{n+1}) = p_1, \forall i = 1, \dots, n] \ \Rightarrow \ [c(p_1, p_{n+1}) \neq p_{n+1}]$ for all $p_1, \dots, p_{n+1} \in X$ and $S_1, \dots, S_{n+1} \in \mathcal{P}(X)$ with $p_i \in S_i$, for all $i = 1, \dots, n+1$. The proof of this is available upon request.

3.7 Conclusion

We propose a new two-period model of anticipation-based and reference-dependent preferences that generalizes Köszegi and Rabin (2006) on the domain of simple lotteries. The theory is based on the concept of anticipatory utility and its effect on utility from actual consumption through the reference point. Furthermore, analogously to Köszegi and Rabin (2006); Kőszegi and Rabin (2007), we define equilibrium concepts that spell out the requirements of choice depending on when the DM actual commits to her decision. In addition, we explicitly show how such a model is capable of rationalizing examples such as information aversion and explicitly lowering ones reference point through strategically anticipating the worst outcome. Sharing the concerns that theories of reference-dependent preferences are difficult to test, we provide characterization and identification results which show that our model is falsifiable based on choice data alone. Our main result show that the observable behavior is characterized by three simple axioms, Expansion, WWARP and NRBC. We additionally show that our main equilibrium concept (PMAE), on the domain of choice, is equivalent to choosing according to a choice procedure, in which the DM first selects a subset of the available alternatives based on a filtering before choosing the most preferred of the remaining alternatives. The equivalence is further underlined by the fact that the first-stage filtering is equivalent to MAE. Finally, we show the extent to which this consideration filter as well as the DM's preference can be revealed through choice data.

3.8 Bibliography

Abeler, J., Falk, A., Goette, L., and Huffman, D. (2011). Reference points and effort provision. *American Economic Review*, 101(2):470–492.

Armouti-Hansen, J. and Kops, C. (2018). This or that? Sequential rationalization of indecisive choice behavior. *Theory and Decision*, 84(4):507–524.

Baillon, A., Bleichrodt, H., and Spinu, V. (2020). Searching for the reference point. *Management Science*, 66(1):93–112.

Bell, D. E. (1985). Disappointment in decision making under uncertainty. *Operations Research*, 33(1):1–27.

Berns, G. S., Chappelow, J., Cekic, M., Zink, C. F., Pagnoni, G., and Martin-Skurski, M. E. (2006). Neurobiological substrates of dread. *Science*, 312(5774):754–758.

Berns, G. S., Laibson, D., and Loewenstein, G. (2007). Intertemporal choice–toward an integrative framework. *Trends in Cognitive Sciences*, 11(11):482–488.

Cherepanov, V., Feddersen, T., and Sandroni, A. (2013). Rationalization. *Theoretical Economics*, 8(3):775–800.

Dekel, E. and Lipman, B. L. (2010). How (not) to do decision theory. *Annu. Rev. Econ.*, 2(1):257–282.

Evers-Kiebooms, G., Nys, K., Harper, P., Zoeteweij, M., Dürr, A., Jacopini, G., Yapijakis, C., and Simpson, S. (2002). Predictive dna-testing for huntington's disease and reproductive decision making: A european collaborative study. *European Journal of Human Genetics*, 10(3):167–176.

Freeman, D. (2016). Preferred personal equilibrium and choice under risk. Technical report, Working Paper.

Freeman, D. J. (2017). Preferred personal equilibrium and simple choices. *Journal of Economic Behavior & Organization*, 143:165–172.

Goeree, J. K. and Holt, C. A. (2001). Ten little treasures of game theory and ten intuitive contradictions. *American Economic Review*, 91(5):1402–1422.

Gul, F. (1991). A theory of disappointment aversion. *Econometrica*, 59(3):667–686.

Harless, D. W. and Camerer, C. F. (1994). The predictive utility of generalized expected utility theories. *Econometrica*, 62(6):1251–1289.

Heidhues, P. and Kőszegi, B. (2008). Competition and price variation when consumers are loss averse. *American Economic Review*, 98(4):1245–68.

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–292.

Kőszegi, B. (2010). Utility from anticipation and personal equilibrium. *Economic Theory*, 44(3):415–444.

Köszegi, B. and Rabin, M. (2006). A model of reference-dependent preferences. *The Quarterly Journal of Economics*, 121(4):1133–1165.

Kőszegi, B. and Rabin, M. (2007). Reference-dependent risk attitudes. *American Economic Review*, 97(4):1047–1073.

Loomes, G. and Sugden, R. (1986). Disappointment and dynamic consistency in choice under uncertainty. *The Review of Economic Studies*, 53(2):271–282.

Manzini, P. and Mariotti, M. (2007). Sequentially rationalizable choice. *American Economic Review*, 97(5):1824 – 1839.

Manzini, P. and Mariotti, M. (2012). Categorize then choose: Boundedly rational choice and welfare. *Journal of the European Economic Association*, 10(5):1141–1165.

Masatlioglu, Y., Nakajima, D., and Ozbay, E. Y. (2012). Revealed attention. *American Economic Review*, 102(5):2183–2205.

Norem, J. K. and Cantor, N. (1986). Defensive pessimism: Harnessing anxiety as motivation. *Journal of personality and social psychology*, 51(6):1208.

Oster, E., Shoulson, I., and Dorsey, E. (2013a). Limited life expectancy, human capital and health investments. *American Economic Review*, 103(5):1977–2002.

Oster, E., Shoulson, I., and Dorsey, E. (2013b). Optimal expectations and limited medical testing: Evidence from huntington disease. *American Economic Review*, 103(2):804–830.

Schmitz, A. and Grillon, C. (2012). Assessing fear and anxiety in humans using the threat of predictable and unpredictable aversive events (the npu-threat test). *Nature Protocols*, 7(3):527–532.

Sen, A. (1971). Choice functions and revealed preference. *The Review of Economic Studies*, 38(3):307–317.

Shalev, J. (2000). Loss aversion equilibrium. *International Journal of Game Theory*, 29(2):269–287.

Spiegler, R. (2008). On two points of view regarding revealed preference and behavioral economics. In Caplin, A. and Schotter, A., editors, *The Foundations of Positive and Normative Economics*, pages 95–115. Oxford University Press, New York.

Sugden, R. (1993). An axiomatic foundation for regret theory. *Journal of Economic Theory*, 60(1):159–180.

Tyson, C. J. (2008). Cognitive constraints, contraction consistency, and the satisficing criterion. *Journal of Economic Theory*, 138(1):51–70.

3.9 Appendix

3.9.1 Proof of Proposition 3.1

Proof: Necessity: Suppose that $C = C(\cdot; U_{\text{MAE}})$. Define $\Gamma : \mathcal{P}(X) \to \mathcal{P}(X)$ as follows: for any $S \in \mathcal{P}(X)$, we define $p \in \Gamma(S)$ if there exists $r_p^* \in R_p$ such that $U(r_p^*, p) \geq U(r_p, q)$, for all $q \in S$ and for all $r_p \in R_p$. Clearly, by this definition, $\Gamma(.)$ satisfies Contraction and Expansion. Hence, by Proposition 3.3, $\Gamma(.)$ is an RD filter.

Sufficiency: Suppose that there exists an RD filter Γ on $\mathcal{P}(X)$ such that $C(S) = \Gamma(S)$ for all $S \in \mathcal{P}(X)$. Then, by Proposition 3.3, C satisfies Contraction and Expansion. Note that it is a well-known fact that contraction and expansion together are logically equivalent to the Weak Axiom of Revealed Preferences (WARP) (Sen, 1971; Tyson, 2008). Standard results then allow us to define a weak preference relation $\succsim$ on X such that $C(S) = \{x \in S | \; x \succsim y, \text{ for all } y \in S\}$ and a function $w : X \to \mathbb{R}$ that represents $\succsim$, i.e., $w(p) \geq w(q)$ if and only if $p \succsim q$. Next, we define a function $U : \Delta \times X \to \mathbb{R}$ such that, for any $p \in X$, $U(r, p) = w(p)$, for all $r \in \Delta$. This implies that $C = C(\cdot; U_{\text{MAE}})$, which establishes the second part of the proof. □

3.9.2 Proof of Proposition 3.2

Proof: Necessity: Suppose that $C = C(\cdot; U_{\text{PMAE}})$. Define a binary relation R over X and a filter Γ over $\mathcal{P}(X)$ as follows: (i) $p \in \Gamma(T)$ if and only if there exists $r_p^* \in R_p$ such that $U(r_p^*, p) \geq U(r_p, q)$, for all $q \in T$ and for all $r_p \in R_p$, and (ii) pRq if and only if there exists $T \in \mathcal{P}(X)$ with $p, q \in \Gamma(T)$ and $r_p^* \in R_p$ such that $U(r_p^*, p) \geq U(r_q, q)$, for all $r_q \in R_q$. Clearly, by this definition, Γ is an RD filter and the strict part of R is acyclic. Let R^* be the transitive closure of R. By Szpilrajn's Theorem, we know that it can be extended to a complete preorder $\succsim$ over X.

Sufficiency: Suppose that there exists a weak preference $\succsim$ and an RD filter Γ such that $C(S) = C(\Gamma(S); \succsim)$, for all $S \in \mathcal{P}(X)$. Since $\succsim$ is complete and transitive, we can define a function $w : X \to \mathbb{R}$ that represents $\succsim$ on X such that $w(p) \geq w(q)$ if and only if $p \succsim q$, for all $p, q \in X$. Since X is finite, we can even define such a representation on $\mathbb{N}$. Next, we use a fixed such $w(\cdot)$ on $\mathbb{N}$ to define $U : \Delta \times X \to \mathbb{R}$ in several steps. First, we define $U(p, p) = w(p)$. Second, for all $p, q \in X$ with $p \succsim q$ and $p \notin \Gamma(\{p, q\})$, we define $U(p, r) = w(r) + (w(p) - w(q) + 1)$, for all r with $r \succsim q$. Third, if $p \succsim q$, $p \notin \Gamma(\{p, q\})$, and the previous step defined $U(r, p)$, for some $r \in R_p$, we furthermore set $U(r, q) = U(r, p) + 1$. Fourth, for all $p \in X$ and all $r \in \Delta$ such that $U(r, p)$ is hitherto undefined, we define $U(r, p) = \min_{q \in X} w(q)$.

To check that $U(\cdot, \cdot)$ thus defined rationalizes $C(\cdot)$, take any $S \in \mathcal{P}(X)$ and consider $C(S) = C(\Gamma(S); \succsim)$. First, we show that all alternatives in $S \setminus \Gamma(S)$ cannot be part of a MAE. To this end, consider any $p \in S$ with $p \notin C(S)$. Then, either (i) $p \succsim q$, for all $q \in S$ and $p \notin \Gamma(S)$, or, (ii) $q \succ p$, for some $q \in S$.

In Case (ii), by our definition of $U(\cdot, \cdot)$ above, $q \succ p$ implies that $U(q, q) > U(p, p)$. Next, suppose, by contradiction, that $U(r_p, p) \geq U(q, q)$, for some $r_p \in R_p$. By our definition of $U(.,.)$ above, this implies that there exists p' with $r_p \succsim q \succ p \succsim p'$ and $r_p \notin \Gamma(\{r_p, p'\})$. But, then, by our definition of $U(.,.)$ above, we have $U(r_p, q) = w(q) + (w(r_p) - w(p') + 1)$ and $U(r_p, p) = w(p) + (w(r_p) - w(p') + 1)$ such that $U(r_p, p) \geq U(r_p, q)$ implies that $w(p) \geq w(q)$ which contradicts $q \succ p$. Hence, $p \notin C(S; U_{\text{PMAE}})$.

In Case (i), by definition of an RD-filter, $p \notin \Gamma(S)$ implies that $p \notin \Gamma(\{p, q\})$, for some $q \in S$. Since $p \succsim q$, our definition of $U(\cdot, \cdot)$ above implies that $U(p, q) > U(p, p)$. Now, suppose $U(r_p, p) > U(p, q)$, for some $r_p \in R_p$. Then, the third step in our definition of $U(\cdot, \cdot)$ above applies such that $U(r_p, q) = U(r_p, p) + 1 > U(r_p, p)$. It follows that $p \notin C(S; U_{\text{PMAE}})$.

Next, we consider any $p \in S$ with $p \in C(S)$. For such p it follows that $p \in \Gamma(S)$ and, by definition of an RD-filter, that $p \in \Gamma(\{p, q\})$, for all $q \in S$. Furthermore, from $p \in C(S)$ it follows that $p \succsim q$ such that $U(p, p) > U(q, q)$, for all $q \in \Gamma(S)$. By our definition of $U(\cdot, \cdot)$ above, it follows that $U(r_p, p) \geq U(r_q, q)$, for all $q \in \Gamma(S)$, $r_q \in R_q$ and $r_p \in R_p$ if and only if $w(p) \geq w(q)$. Hence, by our definition of $w(\cdot)$ above, p is a PMAE if and only if $p \succsim q$, for all $q \in \Gamma(S)$. □

3.9.3 Proof of Proposition 3.3

Proof: We show that Γ is an RD filter if and only if it satisfies Contraction and Expansion

Necessity: Let Γ be an RD filter.

a) Contraction. Let $p \in \Gamma(T)$ and $p \in S \subset T$. Then $\mathcal{C} = \{T\}$ is a cover of S and, by definition of the RD filter, it follows that $p \in \Gamma(S)$, as well.

b) Expansion. Let $p \in \Gamma(S) \cap \Gamma(T)$. Clearly, $\mathcal{C} = \{S, T\}$ is a cover of $S \cup T$. Hence, by definition of the RD filter, it follows that $p \in \Gamma(S \cup T)$, as well.

This establishes necessity of Contraction and Expansion for an RD filter.

Sufficiency. Let Γ be a consideration set mapping that satisfies both Contraction and Expansion. Now, consider any $S \in \mathcal{P}(X)$ and any $p \in S$. Let $\mathcal{C}$ be a cover of S such that $p \in \Gamma(T)$, for all $T \in \mathcal{C}$. Then, Expansion implies that $p \in \Gamma(\bigcup_{T \in \mathcal{C}} T)$. Since $S \subseteq \bigcup_{T \in \mathcal{C}} T$, by Contraction, it follows that $p \in \Gamma(S)$. This establishes our desired conclusion. $\square$

3.9.4 Proof of Theorem 3.1

Proof: Necessity: Let C be an RDC on X, $\succsim$ be a preference relation on X and Γ be an RD filter.

a) Expansion. Let $p \in C(S) \cap C(T)$, for $S, T \in \mathcal{P}(X)$ and some $p \in S \cap T$. For Expansion to hold, we have to show that this implies that $p \in C(S \cup T)$. Clearly, by definition of an RDC, $p \in C(S) \cap C(T)$ implies that $p \in \Gamma(S)$ and $p \in \Gamma(T)$, for if an alternative is chosen from a set, it must be considered in that set. Note that for any $q \in S \cup T$, it holds that either $q \in S$, or, $q \in T$. Hence, since $p \in \Gamma(S)$ and $p \in \Gamma(T)$, the definition of the RD filter implies that $p \in \Gamma(S \cup T)$. Now, suppose, by contradiction, that there exists $q \in \Gamma(S \cup T)$ such that $q \succsim p$ and $\neg[p \succsim q]$, i.e., $q \succ p$. By $q \in \Gamma(S \cup T)$, it follows that $q \in S \cup T$ and that either $q \in S$, or, $q \in T$. W.L.O.G., assume that $q \in S$. Then, by definition of the RD filter, $q \in \Gamma(S \cup T)$ implies that $q \in \Gamma(S)$, as well. But, then, the definition of RDC and $q \succ p$ imply that $p \notin C(S)$ and we have arrived at our desired contradiction. Hence, by completeness of $\succsim$, it follows that $p \succsim q$, for all $q \in \Gamma(S \cup T)$, and, thus, $p \in C(S \cup T)$.

b) WWARP. Let $S, T \in \mathcal{P}(X)$ be such that $q \notin C(\{p, q\}), p \in C(T)$ and $\{p, q\} \subset S \subset T$. For WWARP to hold, we have to show that this implies that $q \notin C(S)$. By definition of an RDC, $p \in C(T)$ implies that $p \in \Gamma(T)$, for if an alternative is chosen from a set, it must be considered in that set. Clearly, by definition of the RD filter, $p \in \Gamma(T)$ implies that $p \in \Gamma(S)$, as well. Now, suppose, by contradiction, that $q \in C(S)$. Clearly, since $\{p, q\} \subseteq \Gamma(S)$, this implies that and $q \succsim p$. By definition of the RD filter, it follows that

$\{p,q\} = \Gamma(p,q)$. But, then, by definition of RDC, $q \succsim p$ implies that $q \in C(\{p,q\})$ and we have arrived at our desired contradiction.

c) NRBC. Let $p_i \in S_i$, $p_i \in C(\{p_i, p_{i+1}\})$, $p_{i+1} \in C(S_i)$, for all $i = 1, \dots, n$, and $p_1 \in C(S_{n+1})$. For NRBC to hold, we have to show that this implies that $p_1 \in C(\{p_1, p_{n+1}\})$. By definition of an RDC, $p_1 \in C(S_{n+1})$ implies that $p_1 \in \Gamma(S_{n+1})$, for if an alternative is chosen from a set, it must be considered in that set. Since $p_{n+1} \in S_{n+1}$ and $p_1 \in \Gamma(S_{n+1})$, by definition of the RD filter, this implies that $p_1 \in \Gamma(\{p_1, p_{n+1}\})$. Analogously, $p_{i+1} \in C(S_i)$ implies that $p_{i+1} \in \Gamma(S_i)$, for all $i = 1, \dots, n$, and, since $p_i \in S_i$, it follows that $p_{i+1} \in \Gamma(p_i, p_{i+1})$, for all $i = 1, \dots, n$. Therefore, by definition of an RDC, $p_i \in C(\{p_i, p_{i+1}\})$ implies that $p_i \succsim p_{i+1}$, for all $i = 1, \dots, n$. Transitivity of $\succsim$ then gives us $p_1 \succsim p_2 \succsim \cdots \succsim p_{n+1}$. Hence, since $p_1 \in \Gamma(\{p_1, p_{n+1}\})$, the definition of an RDC implies that $p_1 \in C(\{p_1, p_{n+1}\})$.

This establishes necessity of the axioms for the representation.

Sufficiency: Suppose that C satisfies the axioms, i.e., Expansion, WWARP, and NRBC. We construct the RD filter Γ and the DM's preferences $\succsim$ on X explicitly. First, define, for any $S \in \mathcal{P}(X)$ with $p \in S$, $p \in \Gamma(S) \;\Leftrightarrow\; p \in C(T)$, for some $T \in \mathcal{P}(X)$ with $S \subseteq T$. Next, for any p, q, define a binary relation R on X by $pRq \;\Leftrightarrow\; p \in C(\{p,q\})$ and $q \in C(S)$, for some $S \in \mathcal{P}(X)$ with $p, q \in S$.

Next let R^* be the transitive closure of R and let P be the asymmetric component of R. It follows that R can be extended to a complete preorder $\succsim$ if and only if it satisfies a variant of acyclicity named only weak cycles (OWC) given by the following condition

$$pR^*q \Rightarrow \neg[qPp]$$

Lemma 3.1 *R on X as defined above satisfies OWC.*

Proof: Let pR^*q for some $p, q \in X$ and suppose, by means of contradiction, that qPp. Since pR^*q, we either have that (i) pRq or (ii) there exists a sequence $(w_m)_{m=1}^M$ in X, such that pRw_1, $w_M Rq$, and for each $m \in \{1, \dots, M-1\}$, $w_m R w_{m+1}$. If (i) is true, then our contradiction follows immediately, so suppose (ii) is true. Next, we establish that the precondition of NRBC for p, $(w_m)_{m=1}^M$, and q holds. To this end, note that pRw_1 implies, by our definition above, that $p \in C(\{p, w_1\})$ and $w_1 \in C(S_1)$, for some $S_1 \in \mathcal{P}(X)$ with $p, w_1 \in S_1$. Next, $w_M Rq$ implies, by our definition above, that $w_M \in C(\{w_M, q\})$ and

$q \in C(S_{M+1})$, for some $S_{M+1} \in \mathcal{P}(X)$ with $w_M, q \in S_{M+1}$. Furthermore, for any $m \in \{1, \ldots, M-1\}$, $w_m R w_{m+1}$ implies, by our definition above, that $w_m \in C(\{w_m, w_{m+1}\})$ and $w_{m+1} \in C(S_{m+1})$, for some $S_{m+1} \in \mathcal{P}(X)$ with $w_m, w_{m+1} \in S_{m+1}$. On the other hand, qPp implies, by our definition above, that $q \in C(\{p,q\})$ and $p \in C(S)$, for some $S \in \mathcal{P}(X)$ with $p, q \in S$. This establishes the precondition of NRBC. On the other hand, by our definition above, qPp also implies that $p \notin C(\{p,q\})$ (as otherwise our definition above would imply that pRq which directly contradicts qPp). This violates NRBC and we have arrived at our desired contradiction. □

We now verify that the objects $(\Gamma, \succsim)$ represent the choice correspondence C on X as an RDC. To that end, pick any $S \in \mathcal{P}(X)$ and let $p \in C(S)$. By our definition of Γ above, it follows that $p \in \Gamma(S)$. Next, we show that there is no alternative $q \in \Gamma(S)$ with $q \succsim p$ and $\neg[p \succsim q]$. Suppose, to the contrary, that there is such a $q \in S$. Then, by our definition of the RD filter above, it follows that there exists a $T \in \mathcal{P}(X)$ such that $S \subset T$ and $q \in C(T)$. Since $\{p,q\} \subset T$, the same definition also implies that $q \in \Gamma(\{p,q\})$ and, thus, by $q \succsim p$ and $\neg[p \succsim q]$, it follows that $p \notin C(\{p,q\})$. But, then $\{p,q\} \subseteq S \subset T$ together with $p \notin C(\{p,q\}), q \in C(T)$ and $p \in C(S)$ violates WWARP and we have arrived at our desired contradiction. It follows that $p \in C(S)$ implies that $p \succsim w$, for all $w \in \Gamma(S)$. This establishes our desired conclusion. □

3.9.5 Proof of Proposition 3.4

Proof: Necessity: Let C be an RDC. Suppose pR^*q does not hold. Then, there exists a weak preference relation $\succsim$ that includes R^* and $q \succsim p$, but $\neg[p \succsim q]$, i.e., it holds that $q \succ p$. By the proof of Theorem 3.1, there exists an RD filter Γ such that $(\Gamma, \succsim)$ represents C. Since $q \succ p$, by definition, p cannot be revealed to be preferred to q.

Sufficiency: We have already shown in Section 3.5 that if pRq, then p is revealed to be weakly preferred to q. Now, consider the case when pR^*q. Since R^* is defined as the transitive closure of R, this implies that there exists a sequence $(w_m)_{m=1}^M$ in X such that pRw_1, w_1Rw_2, $\ldots$, w_MRq. In this case, we know that for any $\succsim$ that is part of a RDC representation, $R \subseteq \succsim$ and, hence, $p \succsim w_1$, $w_1 \succsim w_2$, $\ldots$, $w_M \succsim q$. Further, since $\succsim$ is transitive it follows that $p \succsim q$ and, hence, p is revealed to be preferred to q. □

3.9.6 Proof of Proposition 3.5

Proof: Necessity: Take any $S \in \mathcal{P}(X)$ and any $p \in S$ with $p \notin \Gamma^*(S)$. By the proof of Theorem 3.1, there exists an RD filter Γ with $p \notin \Gamma(S)$ and a preference $\succsim$ such that $(\Gamma, \succsim)$ represents C. Since $p \notin \Gamma(S)$, by definition, p cannot be revealed to receive consideration at S.

Sufficiency: Let $p \in \Gamma^*(S)$. Then, $p \in C(T)$, for some $T \in \mathcal{P}(X)$ with $S \subseteq T$. Clearly, $p \in C(T)$ implies that $p \in \Gamma(T)$. By definition of the RD filter, $S \subseteq T$ and $p \in \Gamma(T)$ imply that $p \in \Gamma(S)$. Then, it follows from the definition of RDC that $p \in \Gamma(S)$ for any Γ that is part of an RDC representation of these choices. $\square$

Chapter **4**

Evaluating the completeness of social preference theories

Abstract. We use machine learning methods as a benchmark for evaluating the predictive capability of simple parameterized social preference theories in a random utility framework. To that end, we use panel data from the lab containing experimental observations of binary dictator games and reciprocity games from Bruhin et al. (2019). To evaluate a given model's predictive capability we apply the concept of a model's *completeness* introduced by Fudenberg et al. (2021), which reveals (i) how large a fraction of the predictable variation of the data a given model captures, and (ii) how large a gain in performance the model brings compared to a naive baseline model. To address the potential remaining patterns in the data that are not captured on the level of representative agent, we conduct the analysis under a mixture model framework allowing for heterogeneity in the estimated parameters.

JEL codes: C52, C53, D11, D12

Keywords: social preferences, dictator games, reciprocity games, theory evaluation, machine learning, random utility models, finite mixture estimation

4.1 Introduction

Laboratory data on choices provides the means to test whether actual decision making matches with proposed theories of choice. In particular, the data makes it possible for us to investigate the extent to which parameterized theories, in their proposed functional form, are able to predict individuals' choices. Such an investigation sheds light on two important points that provide insights on the vary nature of decision making on the considered domain. Firstly, given a theory's included behavioral motives, it allows us to conclude how well the theory is able to predict the choices, compared to how well a theory could have predicted on the considered domain. In turn, this allows us to conclude (i) whether the proposed functional form is optimal and (ii) how much better a theory could perform by considering more complex functional forms. Secondly, it allows us to conclude the extent to which the behavioral motives included in the model matter for decision making. On the domain of other-regarding preferences, such motives might, for instance, be inequity aversion or reciprocity.

In this paper, we address these two points on the domain of other-regarding preferences.[1] In particular, we address them by evaluating simple linear parameterized social preference theories using data on binary dictator games and reciprocity games from Bruhin et al. (2019). The social preference theories are designed in a way that gradually increases the complexity by sequentially adding behavioral motives. Our starting point is a simple linear preference model, in which the decision maker (DM) only cares about her own payoff. By sequentially adding more motives, our end point is a model that includes potentially inequity aversion (or, alternatively, differentiated altruism) and both negative and positive reciprocity.[2]

The insights mentioned above follow from the predictive capability of the models. However, merely looking at the predictive capability of a model does not reveal the whole picture. In particular, when we construct theories, we are potentially not including every potential motive that may influence the choice. Furthermore, the act of choosing might

[1] We use the terms other-regarding preferences and social preferences interchangeably.

[2] To be precise, when we allow the weight that a decision maker assigns to her counterpart utility to depend on their relative payoffs as in Fehr and Schmidt (1999), we will refer to them as follows: If the weight is positive when the DM earns more than her counterpart, we denote this as altruism when ahead, even though one might consider it to be aheadness aversion. Analogously, if the weight is positive when the DM earns less, we will refer to this as altruism when behind, even though it may be a combination of altruism and behindness aversion.

be random on its own. In addition, there may be subjective variations influencing the choice, that we do not account for if we restrict ourselves to the representative agent level. Thus, conditional on the included variables in the theory, we should expect some randomness in choice, leading to less than perfect predictions. It follows that to evaluate the predictive capability of a given model, we need a measure that informs us on how well we could predict, conditional on the included variables used in the models. Such a measure would directly show us the potential improvement, in terms of predictive capability, an alternative formulated theory could bring. Hence, this also allows for the comparison of two models, such that the improvement of including a behavioral motive becomes clear.

In order to conduct this analysis, we first translate the social preference models into prediction rules by subsuming a random utility framework in the same manner as Bruhin et al. (2019). Subsequently, we apply the concept of a model's *completeness* as proposed by Fudenberg et al. (2021). A given parameterized model's completeness is calculated by the improvement in predictive capability the model brings compared to a naive benchmark model, relative to the largest possible improvement in predictive capability in the data. The naive benchmark model in our setting is based on a simple linear model that is stripped of any other-regarding preferences. Hence, this coincides with the predictions we would make if we consider a selfish agent. To calculate the largest possible improvement in prediction, we use machine learning (ML) methods, which allows for a non-parametric and flexible estimation of the predictive patterns in the data.

On the aggregate level, our findings show, that the full linear model that includes all of the considered other-regarding behavioral motives, achieves a relatively high completeness of approximately 82%. Thus, the potential improvements in terms of predictive capability of considering alternative functional forms is quite limited. In addition, the findings indicate that (i) altruism is more important on this domain than reciprocity, (ii) letting altruism depend on whether the DM earns more or less than her counterpart substantially raises the completeness of the model, and (iii) positive reciprocity seems to be slightly more important than negative reciprocity.

We subsequently extend the setting by allowing, in each model, heterogeneity in the parameters as in Bruhin et al. (2019). That is, in each of the models we allow for the existence of several types, each characterized by their own set of parameter values. To evaluate the completeness of a model in this setting, we propose and explore two extensions of the original definition of completeness. The first variant is what we refer to as a model's *within-type completeness.* Here we evaluate the completeness of a model by estimating the completeness within each type that the given model proposes. Specifically, we compare the predictive capability within the type of a given model to that of a ML model, as the estimate of the optimal predictive performance, and to that of a simple model, as the naive benchmark. Besides allowing us to estimate the partial impact of a given behavioral motive, this will allow us to infer (i) whether there is substantial variation in a model's predictive capability across the types, and (ii) whether, for some of the types, a more complex social preference model is needed to fully capture the within-type behavior.

The second variant that we introduce is what we call a model's *unrestricted completeness.* Here we evaluate the predictability of a heterogeneous model by comparing its predictive performance to a fully flexible ML model that uses the subject identifier as a feature. This will provide us with an indication on how well a parametric theory consisting of a parsimonious representation of individuals, in the form of types, predicts compared to a fully flexible non-parametric model that may adjust its predictions to any of the subjects.

Our within-type completeness results on this domain suggest the existence of three types in all of the considered models, with two relatively large ones and one minority type. The behavior of subjects belonging to the first of the large types, which can be characterized by strong other-regarding preferences, seems to be very well predicted by linear social preference models, with completeness estimates ranging between 88% and 93%. The behavior of the second-type subjects is characterized by modest other-regarding preferences. However, the linear social preference theories are only able to achieve a within-type completeness of between 60% and 65%. This indicates that a more complex theory is needed to fully capture this type's behavior. Finally, for the minority type, we find that choices are very random, in the sense that only using the subjects' own payoffs for prediction leads to relative poor predictions. However, due to the type's small size, we do not have enough power to estimate the within-type completeness.

The unrestricted completeness results indicate that a linear social preference model with only three types is able to capture most of the individual variation in the data.

In particular, the completeness estimates range between approximately 85% and 88%. However, we stress that these estimates should be seen as upper bounds of the models' unrestricted completeness, as we cannot claim to have estimated the largest possible improvement in terms of predictability on this expanded feature space. There may exist more complex methods that lead to better predictions.

Our paper contributes to the recent literature on theory evaluation. The most related paper is naturally that of Fudenberg et al. (2021), in which they propose the concept of completeness and evaluate it on the domains of risky choice, initial play in games, and human perception of randomness. Most notably, their findings suggest that, on the aggregate level, cumulative prospect theory (CPT) is 95% complete. In a complementary contribution, Fudenberg et al. (2020) introduce the concept of a model's restrictiveness. This measure is intended to evaluate non-linear paramaterized models' ability to explain real behavior, but exclude artificial infeasible behavior. Specifically, the authors propose a measure to (i) generate artificial data based on a prior distribution, and (ii) evaluate the model's completeness on randomly selected samples of this data. If a model, in general, shows a high completeness on these samples, then a model is unrestrictive. As we only consider linear models, this measure is not directly implementable in our setting. One of the most often used domains in the literature does indeed seem to be that of choice under risk. For instance, Peysakhovich and Naecker (2017) use ML models in the form of regularized regression as a benchmark to evaluate the predictive capability of choice under risk and ambiguity. Specifically, they compare the performance of an expected utility function that admits probability weighting to regularized regression on the domain of risky choice. They do this both for the representative agent and on the individual level. Furthermore, they perform the same analysis for a parameterized second order expected utility model on the domain of ambiguity. Another recent related contribution is that of Fudenberg and Puri (2021). Here completeness is again evaluated on the domain of risky choice. However, in addition to the previous contributions they evaluate the completeness of CPT combined with a parameterized preference for simplicity. Furthermore, this is done in a mixture model framework similar to ours, and to the best of our understanding, their completeness measure coincides with our unrestricted completeness. Another related notable contribution is that of Peterson et al. (2021). Here the authors provide the means to both evaluate proposed theories of risky choice and derive insights to discover new theories by using the by far largest experiment to date. Their findings suggest that proposed theories, such as CPT, perform quite well on limited data and domains. However, on larger data sets, the best performance is achieved by a mixture of theories (MOT) model that, from the context, learns to apply one of two utility functions and one of two probability weighting functions. Other related contributions include, among others, those of Noti et al. (2016), Plonsky et al. (2017) and Plonsky et al. (2019) in which combinations of ML methods and behavioral motives are utilized to predict individual choice. Based on this, our main contribution is the extension of theory evaluation to the domain of social preferences. Naturally, our contribution is limited in (i) the complexity of the models that we consider, and (ii) the range of games, payoffs, and number of counterparts within the games. Thus, our contribution should be seen as a starting point of the evaluation of social preference theories.

The remainder of the paper is organized as follows. The next section describes the setup. In this section, the primitives of our investigation will be defined, the data that we are using will be described, and the social preference models will be presented and translated into parametric models that predict the probability of a decision maker choosing one allocation over the other. Section 4.3 describes our estimation strategy for evaluating the completeness of the models on the aggregate level, as well as for the evaluation of within-type completeness and unrestricted completeness given type heterogeneity. In Section 4.4, we present the findings of our investigation, first on the aggregate level, and subsequently in the heterogeneous setting. Section 4.5 discusses our findings. Finally, Section 4.6 concludes.

4.2 Setup

In this section, we firstly lay out the primitives central to our investigation, based on that of Fudenberg et al. (2020), and present the concept of a parametric model's *completeness*, as proposed by Fudenberg et al. (2021). Secondly, we introduce the data as well as the social preference models that we consider. Afterwards, we show how the social preference models can be translated into parametric models that, given a set of features that constitutes a game, predict the probability that a subject chooses one allocation over the other.

4.2.1 Primitives

Let X be an observable random feature vector from a finite set $\mathcal{X} \subset \mathbb{R}^d$, and Y be a random outcome variable taking values in $\{0,1\}$. Let P denote the joint distribution of (X,Y) and $P_{Y=1|X}$ the conditional probability distribution of $Y=1$ given X. We assume that $P_{Y=1|X}$ is non-degenerate. Our goal is to estimate a mapping $p : \mathcal{X} \to [0,1]$ that enables us to learn the conditional distribution, $P_{Y=1|X}$. Let the set of all possible functions be given by $\mathcal{P}$ and let a function that correctly learns the conditional distribution be given by $p^* \in \mathcal{P}$, i.e. $p^*(x) := P_{Y=1|X=x}$ for all $x \in \mathcal{X}$. We call p^* an optimal mapping. To evaluate the error (or loss) of predicting $p(x)$ given (x,y), we use the negative log-likelihood, which is given by

$$\ell(y, p(x)) = -\left(y \log(p(x)) + (1-y)\log(1-p(x))\right) \tag{4.1}$$

Note that if $y=1$, then Equation (4.1) simplifies to the negative logarithm of the probability $p(x)$ assigns to that event. Hence, the error is smaller the closer $p(x)$ is to 1. Analogously, if $y=0$, then Equation (4.1) simplifies to the negative logarithm of one minus the probability that $p(x)$ assigns to $y=1$. Thus, the error is smaller the closer $p(x)$ is to 0. It is a well-known fact that minimizing the expectation of Equation (4.1), $e_P(\ell(p)) := \mathbb{E}_P[\ell(Y, p(X)]$, is equivalent to minimizing the expected Kullback-Leibler divergence, a measure of the dissimilarity between the estimated and true distribution. Hence the expected loss is minimized by the true conditional distribution. In such a prediction problem, it thus follows that the *irreducible loss* is given by $e_P(\ell(p^*))$. Denote the expected difference (or distance) in loss between any $p, p' \in \mathcal{P}$ by $d_P(p,p')$. It follows immediately that the expected loss of any $p \in \mathcal{P}$ can be formulated as

$$e_P(\ell(p)) = e_P(\ell(p^*)) + d_P(p, p^*) \tag{4.2}$$

Where the first term is the irreducible loss, and hence a lower bound of loss on the considered feature space. The second term tells us how far below the mapping p is of the optimal mapping in terms of predictive performance. Given finite data, the term will, in general, consist of two sources of error. The first source relates to the misspecification of

p in comparison to p^*, that is, the bias. If we impose parametric restrictions, we may not be able to capture all of the predictive patterns. The second source relates to the variance of p. Fitting very flexible functions on finite data may lead to overfitting and may thus introduce variance, in the sense that the estimation would vary substantially based on the finite data set drawn from P. Notice that this also raises a challenge when one wishes to estimate the irreducible loss from finite data. In general, we wish to search $\mathcal{P}$ for the optimal mapping p^*. However, this might require much data. In Section 4.3, we will discuss ways in which we estimate the irreducible loss and when the limited data may force us to choose a suboptimal estimation approach, such that our estimations rather should be viewed as an upper bound of the irreducible loss. Nevertheless, given enough data, having an estimate of the irreducible error informs us how well simple models predict the data and is central to the concept of a parametric model's completeness that we will introduce shortly.

Let $\mathcal{P}_\Theta = \{p_\theta | \theta \in \Theta\}$ be a parametric model, where Θ denotes the parameter space. The specific parametric models that we consider in our application will be defined in Section 4.2.4. For any parametric model, we are interested in a measure that tells us how good it predicts compared to (i) the optimal mapping described above and (ii) a naive baseline predictive mapping. This will inform us on (i) the potential improvement by allowing for more complex interactions and (ii) how much better than a simple mapping the parametric model performs.[3] For this, let $\mathcal{P}_{\Theta_0}$ be a *naive* parametric model. In our setting, as we will show in Section 4.2.4, this will be a simple parametric model with a single parameter. For any parametric model, we assume that $\mathcal{P}_{\Theta_0} \subset \mathcal{P}_\Theta \subset \mathcal{P}$. For the models $\mathcal{P}_\Theta$ and $\mathcal{P}_{\Theta_0}$, denote the optimal model, in terms of lowest expected loss, by p_{θ^*} and $p_{\theta_0^*}$, respectively.[4] The following definition introduces the concept of completeness.

Definition 4.1 (Completeness (Fudenberg et al., 2021)) *Let X and Y be a random feature vector and random outcome variable, respectively, jointly distributed according to P. Furthermore, let $\mathcal{P}_{\Theta_0} \subset \mathcal{P}_\Theta \subset \mathcal{P}$. The completeness of the parametric model $\mathcal{P}_\Theta$ is given by*

$$\kappa_P(\mathcal{P}_\Theta) := \frac{e_P(\ell(p_{\theta_0^*})) - e_P(\ell(p_{\theta^*}))}{e_P(\ell(p_{\theta_0^*})) - e_P(\ell(p^*))} \tag{4.3}$$

[3] For example, how much better a parametric model performs by adding linear altruism compared to one that ignores such behavioral aspects.

[4] Specifically, $\theta^* = \arg\min_{\theta \in \Theta} e_P(\ell(p_\theta))$ and $\theta_0^* = \arg\min_{\theta_0 \in \Theta_0} e_P(\ell(p_{\theta_0}))$.

A parametric model's completeness is thus the ratio of the reduction in expected loss relative to a naive benchmark, compared to the largest possible reduction. Hence, $\kappa(\mathcal{P}_\Theta) \in [0,1]$, where a completeness of 0 implies that the parametric model performs no better than the naive benchmark, and a completeness of 1 implies that the parametric model perform as well as the optimal mapping. Naturally, for $\kappa(\mathcal{P}_\Theta)$ to exist, we need to impose the restriction that $d_P(p_{\theta_0^*}, p^*) > 0$. Specifically, we need to assume that the optimal mapping is strictly more predictive than the naive benchmark, which in most cases can be seen as a non-restrictive assumption.

4.2.2 Data

In our application, we use data from Bruhin et al. (2019) on binary dictator games and reciprocity games. The data consists of two sessions in which the same subjects, who were students at the University of Zürich at the time, faced the same set of dictator games and reciprocity games. We use the data from the first session. Both sessions contain the 174 subjects that participated in both sessions. The subjects made 117 decisions in the active role of Player A. Thus, we have 20,358 observations on subjects in the role of Player A. In addition to the choices, individual characteristics such as age and gender as well as cognitive ability and Big 5 measures were collected by means of a questionnaire. Table 4.5 in the Appendix summarizes these characteristics and measures over the subject.[5]

Of the 117 binary decisions, 39 were dictator game decisions in which subjects in the role of Player A were matched with subjects in the role of Player B, who had no active role. In each of the dictator games, Player A was confronted with two possible allocations, a and b. Each allocation contained a payoff for Player A and a payoff for Player B given by π^A and π^B, respectively. Thus, the allocations a and b were given by (π_a^A, π_a^B) and (π_b^A, π_b^B), respectively. The dictator games were constructed such that 1/3 were games in which $\pi^A > \pi^B$ regardless of whether Player A chose a or b. Analogously, 1/3 were games in which $\pi^A < \pi^B$ regardless of whether Player A chose a or b. Finally, the remaining 1/3 were constructed such that Player A would earn a higher payoff from one of the allocations and Player B would earn a higher payoff from the other. In addition, in each of these three

[5] Note that each subject's payment was based on a show-up fee, a fixed amount for completing the questionnaire and the amounts of the outcomes of three randomly chosen games. The average payment in the first session was 52.5 CHF (see Bruhin et al. (2019) for more information.).

scenarios, games were constructed to vary Player A's cost of altering Player B's payoff in such a way that inequity aversion (or differentiable altruism) parameters were identifiable in the range $[-3, 1]$.

The remaining 78 of the binary decisions were reciprocity games. In a reciprocity game, Player B first decided whether to implement allocation $c = (\pi_c^A, \pi_c^B)$. If she chose to do so, the game ended. If she chose not to do so, Player A, in the second stage, got to decide among allocation a and b. Choices of player A were elicited using the strategy method. The 78 reciprocity games were constructed such that 39 of them were "negative" reciprocity games and the remaining 39 were "positive" reciprocity games. Each of the 39 "negative" and "positive" reciprocity games, respectively, consisted of the same allocation combinations as in the dictator games. A reciprocity game was denoted "negative" if the decision of Player B to not implement c, and thus let Player A choose between a and b, implies that player A would receive a strictly lower payoff in any of the two feasible allocations compared to c. Analogously, a reciprocity game was denoted "positive" if Player A received a strictly higher payoff in any of a and b compared to c. The remaining case in which Player A is better of in one allocation and worse of in another of a and b compared to c were not considered.

4.2.3 Social preference models

In this section, we present the simple social preference models which we will translate into parametric models by adding a random component such that they predict the probability of Player A choosing allocation a over b in any game in Section 4.2.4. The models are capable of capturing other-regarding aspects, such as altruism, inequity aversion, and reciprocity in a simple way. The models will be presented in an ordered mode in which every next step presents added complexity in the form of an additional behavioral aspect. Let u^A and u^B denote the utility of Player A and Player B, respectively. Further, let π^A and π^B denote A's and B's payoff, respectively, from a given allocation.

The first model we specify is one in which Player A lacks any form of other-regarding preferences. Accordingly, her utility is simply given by

$$u_0^A = \pi^A \tag{4.4}$$

Notice that there are no parameters in the utility to be estimated. In our estimations of completeness, the naive benchmark will be based on this model.

Suppose now that Player A's preferences can be described by a simple altruism model. That is, regardless of whether Player B will end up with more or less than Player A in a given allocation, Player B's payoff enters Player A's utility in the same way. Formally, her utility is then given by

$$u_1^A = \pi^A + \gamma_S(\pi^B - \pi^A) \tag{4.5}$$

where γ_S is a parameter that dictates how much Player A cares about Player B's payoff. If $\gamma_S > 0$ then Player A's utility increases in Player B's payoff, ceteris paribus, and she thus exhibits altruism towards Player B. If $\gamma_S < 0$ then Player A always attains a higher utility by decreasing Player B's payoff. If this is the case, then we say that Player A exhibits malice.[6]

The next model we consider is a variant of the inequity aversion model by Fehr and Schmidt (1999). In their model, Player A dislikes outcomes in which she receives a higher payoff than her counterpart and outcomes in which she receives a lower payoff. However, the magnitude to which she dislikes these two situations of inequity may vary. Her utility is defined by

$$u_2^A = \pi^A + (\gamma_D \mathbf{1}_B + \gamma_A \mathbf{1}_A)(\pi^B - \pi^A) \tag{4.6}$$

where $\mathbf{1}_B$ and $\mathbf{1}_A$ are dummy variables taking the value 1 if $\pi^B > \pi^A$ and $\pi^B < \pi^A$, respectively. In our application of the model we do not require that Player A is inequity averse. As such we do not place restrictions on the parameters γ_D and γ_A. If $\gamma_D < 0$ then the parameter denotes Player A's "behindness" aversion or malice when having a lower payoff. On the other hand, if $\gamma_D > 0$ the parameter can be interpreted as the level of altruism when π^A is lower than π^B. Similarly, if $\gamma_A > 0$ then the parameter denotes Player A's "aheadness" aversion or, alternatively, the level of altruism when π^A is higher than π^B. If $\gamma_A < 0$, then the parameter measure A's malice when she is ahead.[7]

[6] Note that specifying the model by $u_1^A = \pi^A + \gamma_S \pi^B$ would lead to the same predictability, but a different parameter estimate. We use our specification as it makes comparing the weight on the the DM's own payoff and that of the counterpart possible.

[7] The reason that we do not restrict the parameters in the sense of Fehr and Schmidt (1999) is that we do not find many subjects that exhibit this type of inequity aversion. As such, the parameter estimate of "behindness" aversion on the aggregate level would go to zero with a significant reduction in the predictive capability. This observation is also the reason why we do not consider other models of inequity aversion as the one by Bolton and Ockenfels (2000).

The remaining models that we consider adds an additional other-regarding aspect, namely reciprocity. Notice that, based on the available data, reciprocity enters binary. As such, in the reciprocity games, we denote $\mathbf{1}_K$ and $\mathbf{1}_U$ as dummy variables that takes the value 1 if Player B's preceding action is deemed "kind" and "unkind", respectively.

Our next model combines simple altruism as defined in Equation (4.5) with both positive and negative reciprocity. Thus, in this simple setting the model can be seen as an implementation of the reciprocal altruism model of Levine (1998).

$$u_3^A = \pi^A + (\gamma_S + \gamma_K \mathbf{1}_K + \gamma_U \mathbf{1}_U)(\pi^B - \pi^A) \tag{4.7}$$

One may interpret γ_K and γ_U as the change Player A's altruism following a kind and unkind action, respectively. Naturally, a result of the estimation with $\gamma_K > 0$ and $\gamma_U < 0$ would be intuitive as the DM would then exhibit positive and negative reciprocity. However, we do not impose that requirement.

We now consider a model that is a direct application of the proposed model of Charness and Rabin (2002) in which the DM's utility form follows an unrestricted inequity aversion model as in Equation (4.6), with the addition that an "unkind" action may affect both of the parameters. In particular, we consider the following functional form

$$u_4^A = \pi^A + (\gamma_D \mathbf{1}_B + \gamma_A \mathbf{1}_A + \gamma_U \mathbf{1}_U)(\pi^B - \pi^A) \tag{4.8}$$

Thus, if $\gamma_U < 0$ then the DM exhibits negative reciprocity and, if she satisfies the requirements of Fehr and Schmidt (1999)-type inequity aversion, then an unkind preceding action of B turns A more "behindness" averse and less "aheadness" averse. Naturally, we do not expect to find behavioral patterns indicating $\gamma_U > 0$.

The final model we consider in this setup is a slight extension of the preceding one in Equation (4.8). Specifically, in addition to "unkind" preceding actions potentially affecting the parameters, we allow "kind" action to do the same as in Bruhin et al. (2019). In this case, the DM's utility is given by

$$u_5^A = \pi^A + (\gamma_D \mathbf{1}_B + \gamma_A \mathbf{1}_A + \gamma_K \mathbf{1}_K + \gamma_U \mathbf{1}_U)(\pi^B - \pi^A) \tag{4.9}$$

Analogously to the previous case, if $\gamma_K > 0$ then the DM exhibits positive reciprocity and if she satisfies the requirements of Fehr and Schmidt (1999) inequity aversion, then a kind preceding action of B turns A less "behindness" averse and more "aheadness" averse.

4.2.4 Parametric models

We now show how the social preference models introduced in Section 4.2.3 can be translated into parametric models, $\mathcal{P}_\Theta$, by adding a random component such that they contain prediction rules for the probability that Player A chooses allocation a over allocation b in a random utility framework, for any game described in Section 4.2.2. We start in a representative agent framework and subsequently expand the setting to allow for heterogeneity. Let the set of games be indexed by $\{1, \ldots, G\}$. A game $g \in \{1, \ldots, G\}$ is uniquely defined by the features $x_g = (x_{g,a}, x_{g,b})$, where $x_{g,m} = (\pi^A_{g,m}, \pi^B_{g,m}, \mathbf{1}_{g,A,m}, \mathbf{1}_{g,B,m}, \mathbf{1}_{g,K}, \mathbf{1}_{g,U})$, for $m \in \{a, b\}$.[8] For each of the social preference models defined in Section 4.2.3, u^A_i for $i \in \{0, \ldots, 5\}$, we assume that Player A's utility of choosing allocation m in game g is the sum of her deterministic utility and noise. Specifically, it is given by

$$u^A_i(x_{g,m}) + \epsilon_i(x_{g,m}) \tag{4.10}$$

Here, $\epsilon_i(x_{g,m})$ is a noise term capturing factors that affect Player A's choice, but is not included in the modelled utility function u^A_i. As in Bruhin et al. (2019), we assume that $\epsilon_i(x_{g,m})$ is Gumbel distributed with scale parameter $\sigma_i > 0$.

For each of the social preference models, let λ_i for $i \in \{0, \ldots, 5\}$ be the set of parameters of model u^A_i and let $y_g = 1$ if the allocation a is chosen by Player A in game g and 0 otherwise. From Train (2009), it follows that the probability of $y_g = 1$ given u^A_i is given by

$$\begin{aligned} p_{\theta_i}(x_g) &= \Pr\left(u^A_i(x_{g,a}) + \epsilon_i(x_{g,a}) > u^A_i(x_{g,b}) + \epsilon_i(x_{g,b})\right) \\ &= \frac{\exp\left\{u^A_i(x_{g,a})/\sigma_i\right\}}{\exp\left\{u^A_i(x_{g,a})/\sigma_i\right\} + \exp\left\{u^A_i(x_{g,b})/\sigma_i\right\}} \end{aligned} \tag{4.11}$$

Where $\theta_i = (\lambda_i, \sigma_i)$ and $\sigma_i > 0$. Notice the role of the scale parameter σ_i. As σ_i decreases the probability of choosing the option with highest deterministic utility increases, so that the choice becomes less noisy. Thus, one may think of $1/\sigma_i$ as the choice sensitivity determining the randomness of choice. Let Θ_i be the set of all possible combinations of

[8] Note that $\mathbf{1}_{g,B,m}$ is here a dummy variable taking the value 1 if Player B's payoff is larger than Player A's in allocation m, in game g and 0 otherwise. Similar explanations hold for the remaining dummy variables. As the inequity dummies contain no additional information when the payoffs themselves are present, we could just as well have removed them from $x_{g,m}$. However, we choose to keep them in to underline the connection between the games and the social preference models.

the parameters.[9] The parametric model $\mathcal{P}_{\Theta_i}$ for $i \in \{0, \ldots, 5\}$ is then given by

$$\mathcal{P}_{\Theta_i} = \{p_{\theta_i} \mid \theta_i \in \Theta_i\} \tag{4.12}$$

To extend our setting to that of heterogeneous agents, we now allow that, for each social preference model $i \in \{1, \ldots, 5\}$, there may exist K types in the population. For this, let $\theta_{i,K} = (\theta_i^1, \ldots, \theta_i^K, \pi_i^1, \ldots, \pi_i^K)$ be the tuple of parameters of model i, where θ_i^k is the parameters of the kth type and $\pi_i^k \geq 0$ is the proportion of that type such that $\sum_{k=1}^K \pi_i^k = 1$. Let $\Theta_{i,K}$ be the set of all possible combinations of the parameters. Based on this, the parametric model $\mathcal{P}_{\Theta_{i,K}}$ is a mixture mode of K types for $i \in \{0, \ldots, 5\}$ and is given by

$$\mathcal{P}_{\Theta_{i,K}} = \left\{ p_{\theta_i,K} = \sum_{k=1}^{K} \pi_i^k p_{\theta_i^k} \mid \theta_{i,K} \in \Theta_{i,K} \right\} \tag{4.13}$$

Where $p_{\theta_i^k}$ follows from Equation (4.11).

4.3 Estimation strategy

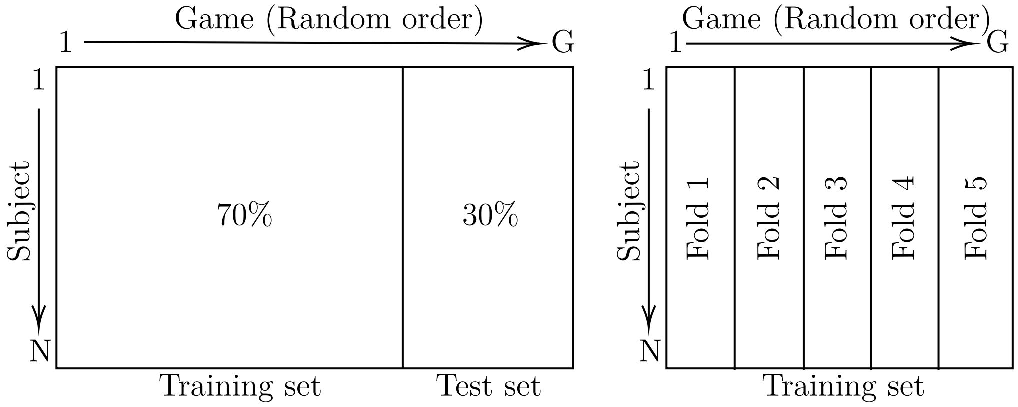
We now describe how we estimate the completeness of the parametric models defined in Section 4.2.4 on two different levels. These two levels vary on whether we allow for heterogeneity in the parameter values of the parametric models. As shown in Section 4.2.1, this estimation involves estimating the irreducible loss. Thus, for each of our applications, we will thoroughly discuss this as well. Before we do so, we briefly describe our data-splitting strategy which allows us to estimate the expected loss of any model.

4.3.1 Data-splitting strategy

We consider a balanced panel $\mathcal{D} = \{(x_{ng}, y_{ng})_{n=1,\ldots,N, g=1,\ldots,G}\}$ consisting of choices of N individuals from G games. Specifically, $x_{ng} = x_g = (x_{g,a}, x_{g,b})$ is the feature vector of game g as defined in Section 4.2.4 and $y_{ng} = 1$ if individual n chose allocation a in game g, and 0 otherwise. In all our applications, our panel will consist of 174 subjects and 117 games, hence $N = 174$ and $G = 117$. Figure 4.1 illustrates our data-splitting strategy, which is similar to that of Peysakhovich and Naecker (2017). In particular, as can be seen

[9] For instance, if $i = 2$, then $\lambda_2 = (\gamma_A, \gamma_D)$, $\theta_2 = (\gamma_A, \gamma_D, \sigma)$ and $\Theta_2 = \{(\gamma_A, \gamma_D, \sigma) \mid \gamma_A, \gamma_D \in \mathbb{R}, \sigma > 0\}$.

Figure 4.1: Data-splitting strategy



Note: The figure on the left shows how the whole panel is randomly split into a training set (70%) and test set (30%), stratifying on the subject level. The figure on the right shows the fold creation in the training set for the 5-fold cross validation procedure.

from the left plot, we randomly split the whole panel data into a training set, consisting of 70% of the observations, $\mathcal{D}_{train}$, and a test set consisting of the remaining 30%, $\mathcal{D}_{test}$, stratifying on the individual level. Specifically, for each individual, we randomly select 70% of the games as train data and the remaining 30% as test data. This means that (i) each individual is present in both the training set and test set and (ii) we approximately have the same number of observations for each individual in both the training set and test set. Notice, however, that since games are randomly chosen on the individual level, we do not necessarily, and most likely will not, observe two individuals with the exact same games in the training set and test set, respectively. We split the data into a train and test set to get an unbiased estimate of the expected loss of a given model. In general, we use $\mathcal{D}_{train}$ to fit our models, and hence estimate the parameters, and use $\mathcal{D}_{test}$ to evaluate out-of-sample predictions. For the simple models on the aggregate level, the difference in estimated loss between in-sample and out-of-sample predictions might be small. However, this will not be the case for the ML models as well as for the parametric models that allow for heterogeneity. Hence, if we would estimate the expected loss based on in-sample predictions, we would get biased estimates heavily favoring the most flexible models.

In addition, when we consider parametric and non-parametric models with hyperparameters, that is, parameters that are not "learned" through model estimation, but rather specified before, we need a method that allows us to choose the optimal of those parameter(s).[10] To perform this model selection we utilize a 5-fold cross validation (CV)

[10] For the ML models that we consider there are usually a variety of hyperparameters to optimize.

technique on the training data. In particular, we randomly partition $\mathcal{D}_{train}$ into five equal sized folds, $\mathcal{D}_{cv}^1, \ldots, \mathcal{D}_{cv}^5$ in the same way that we split the training and test set. The right plot of Figure 4.1 illustrates the folds on the training set. Based on these folds, the CV procedure is as follows: For a given model and for any given combination of hyperparameters, we train and evaluate the model 5 times. Specifically, in iteration $i = 1, \ldots, 5$, we fit the model on all folds except of the ith fold, $\bigcup_{j \neq i} \mathcal{D}_{cv}^j$, and estimate the loss based on the prediction of the model on the ith fold, $\mathcal{D}_{cv}^i$. This results in five loss estimates of which we take the mean as an estimate for model selection.[11]

4.3.2 Aggregate estimations

Note that in all estimations in this section, we ignore the subject identifier. That is, none of our models on the aggregate level make use of the additional information provided by knowing which individual made a given choice. We first show how the parameters and the expected loss of the parametric models are estimated. Afterwards, we describe how we estimate the optimal mapping, its expected loss and the completeness.

4.3.2.1 Parametric models

For the parametric models $\mathcal{P}_{\Theta_i}$, for $i \in \{0, \ldots, 5\}$, our estimation follows two steps in which we wish to (i) estimate the optimal parameters, $\hat{\theta}_i^*$, and (ii) get an estimate of the expected loss of the parametric model given the optimal parameters, $\hat{e}(\ell(p_{\hat{\theta}_i^*}))$. The estimate of (i) follows on the training set, $\mathcal{D}_{train}$ and is given by

$$\hat{\theta}_i^* = \underset{\theta_i \in \Theta_i}{\operatorname{argmin}} \frac{1}{|\mathcal{D}_{train}|} \sum_{(x,y) \in \mathcal{D}_{train}} \ell(p_{\theta_i}(x), y), \quad \text{for } i \in \{0, \ldots, 5\} \tag{4.14}$$

Thus, the optimal parameters of a given parametric model are the ones in the parameter space that results in the lowest average negative log-likelihood on the training set, $\mathcal{D}_{train}$. Based on our estimate of the optimal parameters for any given parametric model, we

Additionally, we view the optimal number of types K in heterogeneous parametric models as a hyperparameter.

[11] Estimating the expected loss and choosing the optimal hyperparameter(s) on the same test set may lead to biased results. See Friedman et al. (2009) for a thorough treatment on model selection and assessment.

estimate (ii) on the test set, $\mathcal{D}_{test}$, as follows

$$\hat{e}(\ell(p_{\hat{\theta}_i^*})) = \frac{1}{|\mathcal{D}_{test}|} \sum_{(x,y)\in\mathcal{D}_{test}} \ell(p_{\hat{\theta}_i^*}(x), y), \quad \text{for } i \in \{0, \dots, 5\} \tag{4.15}$$

Such that our estimate of the expected loss of parametric model i is the average negative log-likelihood on the test set, $\mathcal{D}_{test}$, from the model's predictions based on the optimal parameters estimated on the train set, $\mathcal{D}_{train}$. This procedure provides us with estimates of the expected error of all the parametric models, including the naive benchmark (i.e. for p_{θ_0}).

4.3.2.2 Completeness

We now turn to our estimation strategy of the optimal mapping, $\hat{p}^* \in \mathcal{P}$, and its expected loss, $\hat{e}(\ell(\hat{p}^*))$, which will serve as our estimate of the irreducible loss. To estimate it, we employ non-parametric and ML algorithms that are highly flexible.

The first algorithm that we consider is a *table lookup algorithm*, p_{TL}. In this simple algorithm, the optimal mapping is estimated as follows: For each game $g \in \{1, \dots, G\}$, let $p_{TL}(x_g)$ be the relative frequency with which allocation a was chosen in the data in that game. Thus, the table lookup algorithm can be seen as a non-parametric estimation of the conditional probability distribution. As such, the algorithm converges to the conditional probability distribution asymptotically, but may be suboptimal on finite data. Naturally, the estimation of the relative frequency is performed on the training set, $\mathcal{D}_{train}$, and the predictions are evaluated on the test set, $\mathcal{D}_{test}$.[12]

The remaining two algorithms that we consider are so-called ensemble ML methods, that may improve over the table lookup algorithm on finite data. That they are ensemble methods refer to the fact that they each consist of a collection of simple prediction rules, but the way in which the decision rules are aggregated to arrive at a prediction defers between the two models. In our application, we consider a *random forest* classifier, p_{RF}, and a *gradient boosting* classifier, p_{GB}, which both contain an ensemble of decision trees (see Friedman et al. (2009)).

[12] Fudenberg et al. (2021) show that the table lookup algorithm slightly outperforms that of a variant of a random forest in three applications. In their setting it is, however, unclear whether and to which extent hyperparameter optimization is performed. Thus, it might be that ML methods outperforms the table lookup algorithm if model selection is applied.

The random forest is an ensemble of decision trees, generally trained via the bagging (bootstrap aggregating) method. This means that each decision tree in the ensemble is fitted on a sample drawn from the training set, $\mathcal{D}_{train}$, with replacement. Out-of-sample predictions on the test data, $\mathcal{D}_{test}$, are then made by averaging each decision tree's prediction in the ensemble. When using such a method, it is important to first find the optimal hyperparameters.[13] To estimate the optimal hyperparameters we utilize the previously described 5-fold CV procedure. That is, we first find the optimal hyperparameters in the 5-fold CV procedure, then we fit the optimal random forest classifier on the training set, $\mathcal{D}_{train}$, and finally, we estimate the expected error by evaluating its predictions on the test set, $\mathcal{D}_{test}$. Analogously to the random forest, the gradient boosting classifier contains an ensemble of decision trees. However, here we consider a sequence of trees in which each tree's objective is to improve the prediction of the preceding ones. Thus, the trees in the ensemble will be interdependent to a much higher degree than in the random forest. As is standard, we use a learning rate hyperparameter to control how much weight to attach to a new tree in the sequence. Thus, here again, it is important to find the optimal hyperparameters, for which we follow the same method as in the random forest.[14]

From all of the above, our estimate of parametric model $\mathcal{P}_{\Theta_i}$'s completeness, $\hat{\kappa}(\mathcal{P}_{\Theta_i})$, for $i \in \{0, \ldots, 5\}$ is given by

$$\hat{\kappa}(\mathcal{P}_{\Theta_i}) = \frac{\hat{e}(\ell(p_{\hat{\theta}_0^*})) - \hat{e}(\ell(p_{\hat{\theta}_i^*}))}{\hat{e}(\ell(p_{\hat{\theta}_0^*})) - \hat{e}(\ell(\hat{p}^*))} \tag{4.16}$$

Where $\hat{e}(\ell(\hat{p}^*)) = \min\{\hat{e}(\ell(\hat{p}^*_{TL})), \hat{e}(\ell(\hat{p}^*_{RF})), \hat{e}(\ell(\hat{p}^*_{GB}))\}$.

4.3.3 Heterogeneous estimations

The original definition of completeness refers to the aggregate level. To evaluate a parametric model's completeness on the heterogeneous level, we consider two variants which both offer insights into the predictive capability of the model. The first variant is the

[13] Important hyperparameters to consider here are (i) the number of trees in our ensemble, (ii) the depth (or size) allowed for each of our trees and the minimum number of observations required to make a split in the decision tree, and (iii) the number of features randomly available to each tree when making a split.

[14] Important hyperparameters for the gradient boosting classifier are (i) number of trees in our ensemble, (ii) the depth allowed for each of our trees and the minimum number of observations required to make a split, since we want to have low complexity trees with high bias, (iii) the learning rate because we do not want to over-adjust predictions based on a single new tree.

within-type completeness. Here we estimate a parametric model's completeness based on the partitioning of subjects into K types determined in the estimation of the model. Based on this, estimates of within-type completeness follows from comparing the predictive performance of a parametric model within each type that it defines to that of a naive benchmark fitted within each of the type-dependent partitions and to that of a ML model that uses the type-partitioning as a feature. Intuitively, we perform this estimation because a parametric model of K types may perform well across types, but exhibit substantial variation within types in terms of predictive capability. Thus having such an estimate will reveal how well each type can be summarized by the given parametric model that we consider, and whether, for some of the types, a more complex social preference model is needed to fully capture their behavior. We thus see this approach as the natural extension of the definition of completeness to the heterogeneous setting.[15] The next variant of completeness that we consider is what we call the unrestricted completeness. Here we evaluate the completeness of the heterogeneous parametric models by comparing their performance to a fully flexible ML model that has information on the subject identifier in each observation. This will provide an indication of how well a given parametric model, with a parsimonious representation of subjects in the form of types, performs compared to a non-parametric model that may adjust its predictions to any of the subjects. In the following, we first show how we estimate the parametric models, and their expected loss, on the heterogeneous level. Afterwards we address the estimation of the optimal mapping and the irreducible loss in this expanded feature space based on the two variants of completeness that we wish to estimate.

4.3.3.1 Parametric models

In order to estimate the optimal parametric models allowing for K types, we treat it as a "missing" data problem, in which each subject $n \in \{1, \dots, N\}$ belong to a type $k \in \{1, \dots, K\}$, but that this type membership in unobservable, as in Bruhin et al. (2019).[16] For this, let $\mathcal{D}_{train_n}$ be the set of observations in the training set that involves

[15] Note that Fudenberg et al. (2021) proposes the use of a clustering algorithm to assign subjects to types. This will naturally result in model-independent type assignment. However, we find that the optimal way of assigning subjects to types should come through the given parametric model as the type assignments should depend on the parameter estimates themselves.

[16] See McLachlan et al. (2019) for a recent review of finite mixture models.

subject n.[17] It follows that subject n's contribution to the likelihood conditional on being type k in model i can be stated by the following

$$\mathcal{L}(p_{\theta_i^k}; n) = \prod_{(x,y)\in\mathcal{D}_{train_n}} p_{\theta_i^k}(x)^y \times (1 - p_{\theta_i^k}(x))^{1-y} \tag{4.17}$$

If we were searching for the optimal parameters on the individual level, i.e., when $K = N$, we would directly choose the parameters that minimize the average negative logarithm of Equation (4.17) for each subject n. However, we are interested in an estimation of a parsimonious representation involving $K \ll N$ types. For this, denote subject n's total likelihood contribution across types by $\sum_{k=1}^{K} \pi_i^k \mathcal{L}(p_{\theta_i^k}; n)$, where π_i^k is the proportion of type k in model i. Based on this, for a given number of types K, the estimated optimal parameters of model $i \in \{1, \ldots, 5\}$ are defined as follows

$$\hat{\theta}^*_{i,K} = \underset{\theta_{i,K}\in\Theta_{i,K}}{\operatorname{argmax}} \sum_{n=1}^{N} \log \left(\sum_{k=1}^{K} \pi_i^k \mathcal{L}(p_{\theta_i^k}; n) \right) \tag{4.18}$$

As the objective function, in general, is not well-behaved, finding the optimal parameters is not a trivial task. However, estimations can be achieved by utilizing the iterative expectation maximization (EM) algorithm (Dempster et al., 1977).[18] An additional upside of the estimation is that the posterior probability of type assignment for each individual can be calculated by Bayesian updating. Hence, given K and the estimated optimal parameters $\hat{\theta}^*_{i,K}$ for model i, the estimated probability that subject n belongs to type k is given by

$$\hat{\tau}_{i,n}^k = \frac{\hat{\pi}_i^k \mathcal{L}(p_{\hat{\theta}_i^k}; n)}{\sum_{j=1}^{K} \hat{\pi}_i^j \mathcal{L}(p_{\hat{\theta}_i^j}; n)} \tag{4.19}$$

We will use these probabilities to classify each subject into types and make out-of-sample predictions on the test set, $\mathcal{D}_{test}$. Specifically, denote by k_n the type k in which subject n most likely belongs, for $n = 1, \ldots, N$, based on estimations on the training set, $\mathcal{D}_{train}$. Based on this, our estimate of the overall expected loss of parametric model $\mathcal{P}_{\Theta_{i,K}}$ is given by

$$\hat{e}(\ell(p_{\hat{\theta}^*_{i,K}})) = \frac{1}{|\mathcal{D}_{test}|} \sum_{(x_{ng},y_{ng})\in\mathcal{D}_{test}} \ell\left(p_{\hat{\theta}_i^{k_n}}(x_{ng}), y_{ng}\right) \tag{4.20}$$

It thus follows that for any subject n, we use the estimated parameters of the type in which she most likely belongs to make predictions in each game that she encounters.

[17] Formally, for $n \in \{1, \ldots, N\}$, $\mathcal{D}_{train_n} = \{(x_{ng}, y_{ng})_{g=1,\ldots,G} | (x_{ng}, y_{ng}) \in \mathcal{D}_{train}\}$.

[18] As problems may be encountered when fitting a finite mixture model with a high number of components, we only consider $K \leq 10$.

This methodology provides us with expected loss estimates for all parametric models $i \in \{1, \ldots, 5\}$ and for a given K. As a parametric model of K types may perform well across types, but exhibit substantial variation within types in terms of predictive capability, we also estimate the within-type expected loss of each type $k \in \{1, \ldots, K\}$. Such an estimation is analogous to the one given in Equation (4.20), following a partitioning of the test set into K sets, $\mathcal{D}_{test_1}, \ldots, \mathcal{D}_{test_K}$, each containing the observations of the individuals who belong to that type based on the estimated type membership probability. That is, the parametric model $\mathcal{P}_{\Theta_{i,K}}$'s expected loss within type k is defined as

$$\hat{e}_k(\ell(p_{\hat{\theta}_i^k})) = \frac{1}{|\mathcal{D}_{test_k}|} \sum_{(x,y) \in \mathcal{D}_{test_k}} \ell\left(p_{\hat{\theta}_i^k}(x), y\right) \tag{4.21}$$

Notice that the number of types K is not "learned" in the fitting stage. As mentioned, we treat K as a hyperparameter and once again utilize the 5-fold CV procedure. This will result in five CV loss estimates for each K that we consider. To find the optimal number of types for each parametric model, we apply the "one-standard-error" rule.[19] Specifically, let $\tilde{K}$ be the number of types in which the parametric model reaches its minimum of the average CV loss across the 5 folds for all K's that we consider and let K^* be the smallest K for which the average CV loss is within one standard error from that of $\tilde{K}$. If that is the case, then we choose K^* as the optimal number of types for model i.[20] Our rationale for applying this selection criterion is a combination of two reasons. Firstly, increasing the number of types increases the dimensionality of the problem and, therefore, increases the variance of our estimations. Thus, the average loss over the CV iterations provides a noisier indication of the predictive capability of the model. Based on this, we prefer a model with a smaller number of types given its estimation is within a reasonable range of the noise. Secondly, our reasoning is also based on an intrinsic preference for parsimony

[19] The "one-standard-error" rule is commonly applied in optimization problems involving a single hyperparameter, such as regularized regression. See, for instance, Friedman et al. (2009).

[20] Estimating the "optimal" number of types K in a mixture model is non-trivial, and there exists a variety of methods in order to do so. In their application, Bruhin et al. (2019) apply the normalized entropy criterion (NEC). The NEC is defined to favor models with a "clean" classification of types in the sense that the probability of a given subject belonging to a given type is either close to zero or one. However, as NEC is undefined for $K = 1$, it is not possible to determine whether a a model with $K > 1$ or $K = 1$ is more suitable based on this criterion. Using CV techniques to determine the optimal K is not a new approach. Smyth (2000), for instance, reports reasonable results by utilizing a CV variant. Finally, other criteria such as the Akaike information criterion (AIC) or the Bayesian information criterion (BIC) have also commonly been applied (see Peel and MacLahlan (2000)). However, depending on whether some regularity conditions, which are hard to check, are satisfied, both of these may suffer from their own set of issues in determining the best K.

in the economic theories. In general, we would prefer that any given economic theory can capture the population with a smaller number of types that are distinct from each other compared to a larger number of types in which type parameters vary only slightly without an economically significant difference.

4.3.3.2 Within-type completeness

We now describe our estimation strategy of the within-type completeness. To estimate the optimal mapping in this setting, we use the ML method in Section 4.3.2 that proved to be the best in predicting choices on the aggregate level. However, in addition to the features that the model could use on the aggregate level, we expand the feature space by adding a type indicator. That is, for each observation, there is also a type indication available to the ML model specifying to which type the individual who made the choice belong according to the given parametric model to which we are comparing. As type membership and the number of types potentially varies between the parametric models, we estimate the ML model separately for each of the parametric models that we consider. Naturally, the hyperparameters for each of the ML models are optimized using the 5-fold CV procedure. This will result in a distinct optimal mapping, $\hat{p}_i^*$, for each of the parametric models. Following this, we estimate the within-type expected loss of $\hat{p}_i^*$ for parametric model i by a partitioning of the test set into into K sets, $\mathcal{D}_{test_1}, \ldots, \mathcal{D}_{test_K}$, each containing the observations of the individuals who belong to that type based on the parametric model. In turn, based on the out-of-sample predictions of $\hat{p}_i^*$ on each of these partitions, this will gives us K within-type expected loss estimates, $\hat{e}_1(\ell(\hat{p}_i^*)), \ldots, \hat{e}_K(\ell(\hat{p}_i^*))$. To estimate the naive benchmark, we fit p_{θ_0} separately on K partitions of the training set, $\mathcal{D}_{train_1}, \ldots, \mathcal{D}_{train_K}$, defined in the same manner as the partitions on the test set. We then once again estimate the within-type expected loss on each partition of the test set, resulting in K within-type naive expected loss estimates, $\hat{e}_1(\ell(p_{\hat{\theta}_0^*})), \ldots, \hat{e}_K(\ell(p_{\hat{\theta}_0^*}))$ for each parametric model. Our estimation of the within-type completeness of parametric model $\mathcal{P}_{\Theta_i,K}$ is then a set of completeness estimates of each type and is given as follows

$$\hat{\kappa}(\mathcal{P}_{\Theta_i,K}) = \left\{ \frac{\hat{e}_k(\ell(p_{\hat{\theta}_0^*})) - \hat{e}_k(\ell(p_{\hat{\theta}_i^k}))}{\hat{e}_k(\ell(p_{\hat{\theta}_0^*})) - \hat{e}_k(\ell(\hat{p}_i^*))}, \text{ for } k = 1, \ldots, K \right\} \tag{4.22}$$

4.3.3.3 Unrestricted completeness

Finally, we can describe our strategy for estimating the unconditional completeness. To estimate the optimal mapping, $\hat{p}^* \in \mathcal{P}$, in this setting, we consider four different variants of the ML method that proved to be the best on the aggregate level. In the first variant, we let the ML method use the subject identifier directly as a feature. Thus, the model is fitted on each observation of the training data containing all of the available information, and we once again optimize the hyperparameters using the 5-fold CV procedure. We denote the optimal model in this variant p^*_{ind}. As using the subject identifier directly may introduce substantial variance due to the relative limited number of observations for each individual, we also consider three additional variants that utilize a clustering algorithm as a preprocessing step, which clusters the subjects into a pre-specified number of groups based on their respective vector of choices over games in the training set, $\mathcal{D}_{train}$. In particular, we consider variants of K-means clustering, hierarchical clustering, and a division of subjects based on a Bernoulli mixture model.[21] We treat the number of clusters in each of these algorithms as a hyperparameter. Thus, the optimal number of clusters is determined jointly with the other hyperparameters of the ML method in the 5-fold CV procedure.[22] We denote the optimal models in these three variants p^*_{km}, p^*_{hc} and p^*_{ber}, respectively. Our estimation of the unconditional completeness of $\mathcal{P}_{\Theta_i,K}$ is then given as follows

$$\hat{\kappa}(\mathcal{P}_{\Theta_i,K}) = \frac{\hat{e}(\ell(p_{\hat{\theta}^*_0})) - \hat{e}(\ell(p_{\hat{\theta}^*_{i,K}}))}{\hat{e}(\ell(p_{\hat{\theta}^*_0})) - \hat{e}(\ell(\hat{p}^*_j))} \tag{4.23}$$

Where $\hat{e}(\ell(\hat{p}^*_j)) = \min\{\hat{e}(\ell(\hat{p}^*_{ind})), \hat{e}(\ell(\hat{p}^*_{km})), \hat{e}(\ell(\hat{p}^*_{hc})), \hat{e}(\ell(\hat{p}^*_{ber}))\}$. Notice that the naive benchmark in Equation (4.23) is the same as the one used for completeness on the aggregate level. Thus, the unrestricted completeness tells us (i) how much a parsimonious heterogeneous representation of a given parametric model improves over a simple naive representation of the subjects on a representative agent level, and (ii) how close it is, in terms of predictive ability, to that of the optimal mapping, that may capture any form

[21] For K-means clustering, we consider a variant similar to that proposed by Chi et al. (2016) allowing for clustering in the presence of missing data, after standardization. The hierachical clustering that we implement follows a standard bottom-up approach in which each subject initially is assigned her own cluster.

[22] Notice that we do not apply the "one-standard-error" rule when we optimize the hyperparameters of the ML models. The reason for this is that we have multiple hyperparameters to optimize. It is therefore not clear which of two potential candidate vector of hyperparameters leads to a more parsimonious model.

of heterogeneity in the data. However, at this point, we would like to clarify that the resulting estimation should be seen as a upper bound of the completeness of the model. There may, in fact, exist methods of clustering the subjects that could result in better performance than what we see here.

4.4 Results

We now present the results of our investigation. We will first show the results of the parametric models' completeness on the aggregate level. Afterwards, we will present the completeness estimations of the models on the heterogeneous level. Specifically, for the optimal heterogeneous parametric models we will investigate the models' within-type completeness. Such estimates will inform us on the potential improvement by extending the model for a given subset of the subjects. Following, we will present the heterogeneous parametric models unrestricted completeness.

4.4.1 Representative agent

Table 4.6 in the Appendix shows the results of the estimation of our non-parametric and ML models. Specifically, the table shows (i) the average CV loss of the ML model with its optimal hyperparameters and (ii) the expected loss estimate of all the models. We see that the gradient boosting classifier $\hat{p}^*_{GB}$, has the lowest expected loss estimate, slightly outperforming that of the random forest, $\hat{p}^*_{RF}$. It is also apparent that the expected loss estimate of the table lookup algorithm, $\hat{p}^*_{TL}$, is substantially higher than that of the other models. Hence, in contrast to the applications in Fudenberg et al. (2021), it appears that that we do not have enough observation such that the algorithm approaches the conditional distribution to a high enough degree. Having found the irreducible loss estimate, $\hat{p}^*$, namely the expected loss estimates of the gradient boosting classifier, $\hat{p}^*_{GB}$, we now present the completeness and parameter estimates of the parametric models on the aggregate level. Table 4.1 presents these, where the completeness estimates of the naive benchmark, $p_{\hat{\theta}^*_0}$, and the optimal mapping, $\hat{p}^*$, are 0% and 100% by definition, respectively.

As can be seen, adding a single altruism parameter, $p_{\hat{\theta}^*_1}$, raises the completeness es-

Table 4.1: Parameter estimates and completeness of parametric models on the aggregate level

Model	$1/\hat{\sigma}$	$\hat{\gamma_S}$	$\hat{\gamma_A}$	$\hat{\gamma_D}$	$\hat{\gamma_K}$	$\hat{\gamma_U}$	$\hat{e}(\ell(\cdot))$	$\hat{\kappa}(\cdot)$
$p_{\hat{\theta}_0^*}$	0.0118^{***}	–	–	–	–	–	0.3862	0%
	(0.0002)							
$p_{\hat{\theta}_1^*}$	0.0128^{***}	0.1660^{***}	–	–	–	–	0.3555	59.41%
	(0.0007)	(0.0161)						
$p_{\hat{\theta}_2^*}$	0.0130^{***}	–	0.2573^{***}	0.0696^{***}	–	–	0.3482	73.40%
	(0.0007)		(0.0196)	(0.0169)				
$p_{\hat{\theta}_3^*}$	0.0129^{***}	0.1568^{***}	–	–	0.0722^{***}	-0.0477^{***}	0.3512	67.66%
	(0.0007)	(0.0172)			(0.0162)	(0.0130)		
$p_{\hat{\theta}_4^*}$	0.0131^{***}	–	0.2849^{***}	0.0970^{***}	–	-0.0845^{***}	0.3454	78.92%
	(0.0007)		(0.0203)	(0.0173)	–	(0.0121)		
$p_{\hat{\theta}_5^*}$	0.0131^{***}	–	0.2483^{***}	0.0602^{***}	0.0726^{***}	-0.0479^{***}	0.3439	81.71%
	(0.0007)		(0.0205)	(0.0181)	(0.0162)	(0.0131)		
$\hat{p}^*$	–	–	–	–	–	–	0.3345	100%
Number of subjects: 174 Number of observations: 20,358								

Note: $p_{\hat{\theta}_0^*}, p_{\hat{\theta}_1^*}, p_{\hat{\theta}_2^*}, p_{\hat{\theta}_3^*}, p_{\hat{\theta}_4^*}$ and $p_{\hat{\theta}_5^*}$ are parametric models as defined in Section 4.2.4. $\hat{p}^*$ is a gradient boosting classifier. $\hat{e}(\ell(\cdot))$ is the average negative log likelihood on the test set. $\hat{\kappa}(\cdot)$ is the estimated completeness on the test set. Standard errors clustered on the individual level in parentheses. Significance levels; ${}^*p < 0.1,{}^{**}p < 0.05,{}^{***}p < 0.01$.

timate up to 59.41%. The altruism parameter estimate, $\hat{\gamma}_S$, in this simple model is statistically significant at the 1%-level and of non-negligible magnitude. In particular, the representative agent is willing to give up roughly 17 cents to raise her counterpart's payoff by one dollar. The estimated choice sensitivity, $1/\hat{\sigma}$, also increases by adding the additional parameter compared to the naive benchmark model, $p_{\hat{\theta}_0^*}$. Thus, the agent's choice becomes less random with its inclusion compared to the naive benchmark.

Next, letting the altruism parameter depend on whether the agent earns a higher or lower payoff than her counterpart, $p_{\hat{\theta}_2^*}$, raises the completeness estimate by approximately 14 percentage points to 73.40% compared to the model with a single altruism parameter, $p_{\hat{\theta}_1^*}$. Both altruism parameter estimates are positive and statistically significant at the 1%-level, such that we do not find evidence for subjects being behindness averse on the aggregate level.[23] The point estimate of altruism when the agent earns less than her counterpart, $\hat{\gamma}_D$ is relatively small, indicated a willingness to give up approximately 7 cents to increase the counterpart's payoff by one dollar. On the other hand, we observe

[23] To be precise, $\hat{\gamma}_D$ may capture altruism and behindness aversion at the same time, with altruism being the stronger of the two motives.

a substantial point estimate of the altruism parameter when the agent is ahead, $\hat{\gamma}_A$. In particular, this shows a willingness to give up approximately 26 cents to increase the counterpart's payoff by one dollar. In turn, it follows that the point estimate of altruism, $\hat{\gamma}_S$, in model $p_{\hat{\theta}_1^*}$ is roughly the mean of the point estimates of altruism, $\hat{\gamma}_A$ and $\hat{\gamma}_D$, in model $p_{\hat{\theta}_2^*}$. Finally, notice that the choice sensitivity increases even further by letting altruism depend on whether the agent earns more or less than her counterpart.

The model that uses a single altruism parameter, but adds both negative and positive reciprocity, $p_{\hat{\theta}_3^*}$, achieves a completeness estimate of 67.66%. Hence, not allowing for differentiated altruism reduces the completeness, even when including reciprocal concerns. In particular, the completeness is reduced by approximately 6 percentage points compared to $p_{\hat{\theta}_2^*}$. In addition, we see that the choice sensitivity slightly decreases, whereas the parameter estimate of altruism, $\hat{\gamma}_S$, is similar to that in $p_{\hat{\theta}_1^*}$. We also see that the parameter estimates of positive and negative reciprocity have the expected signs, are statistically significant at the 1%-level and have relative small effect sizes (0.0722 and -0.0477, respectively), with positive reciprocity having a slightly larger impact on utility than that of negative reciprocity. This provides us with a first indication that differentiated altruism is the most important behavioral motive on this domain.[24]

The next model that incorporates all behavioral motives except of positive reciprocity, $p_{\hat{\theta}_4^*}$, substantially improves in the estimated completeness compared to the previous one. The estimated completeness is 78.92% and it therefore also improves over $p_{\hat{\theta}_2^*}$. In particular, on this domain, adding negative reciprocity to a model of differentiated altruism increases completeness by approximately 6 percentage points. In terms of parameter estimates, we see that (i) all are significant at the 1%-level, (ii) both altruism parameter estimates, $\hat{\gamma}_A$ and $\hat{\gamma}_D$, are higher than that of $p_{\hat{\theta}_2^*}$, and (iii) the parameter estimate of negative reciprocity, $\hat{\gamma}_U$, is lower than that of $p_{\hat{\theta}_3^*}$. This suggests that positive reciprocity plays a role on this domain, such that the altruism parameters absorbs this effect. In turn the negative reciprocity parameter needs to be lower to compensate for this. Additionally, we see that the choice sensitivity increases slightly.

Finally, we consider the model that incorporates all motives, $p_{\hat{\theta}_5^*}$.[25] We once again

[24] In general, this might not be the case in richer domains where reciprocity enters in a non-binary way.

[25] Note that our parameter estimates of this model are similar to but distinct from that of Bruhin et al.

see an increase in the estimated completeness. The estimated completeness is 81.71% suggesting that the partial impact of positive reciprocity is approximately 3 percentage points compared to the previous model. We see that all parameter estimates are significant at the 1%-level and that the choice sensitivity is similar to that in the previous model. Furthermore, we see that the parameter estimates of altruism, $\hat{\gamma}_A$ and $\hat{\gamma}_D$, are similar to that of $p_{\hat{\theta}_2^*}$ and that the reciprocity parameter estimates, $\hat{\gamma}_K$ and $\hat{\gamma}_U$, are similar to that of $p_{\hat{\theta}_3^*}$. Thus, we find that (i) letting altruism differentiate depending on the payoff allocations raises completeness substantially and hence captures the choices of representative agent significantly more accurately and (ii) the potential improvement of considering a more flexible functional form is quite limited as the model that includes both differentiated altruism and reciprocity linearly, $p_{\hat{\theta}_5^*}$, captures more than 4/5 of the predictable variation in the data, relative to the naive benchmark.[26] Based on these results, we exclude $p_{\hat{\theta}_1^*}$ and $p_{\hat{\theta}_3^*}$ from the heterogeneous analysis.

4.4.2 Heterogeneity

We now consider the estimation of completeness on the heterogeneous level. We first show the estimations of the within-type completeness of the parametric models, followed by the parameter estimates of the models. Afterwards, we present the estimations of the unrestricted completeness.

Before we present the completeness estimations, we briefly address the selection of the optimal number of types for each of the parametric models that we consider. Figure 4.2 in the Appendix shows the CV loss estimates of the heterogeneous parametric models $p_{\hat{\theta}_{2,K}}, p_{\hat{\theta}_{4,K}}$ and $p_{\hat{\theta}_{5,K}}$ for $K = 1, \ldots, 10$, where K is the number of types. For all of the three models, we see a significant decrease in the CV loss going from $K = 1$ to $K = 3$. After this initial decrease we see a flattening of the curve, in which the estimated loss either increases or decreases slightly in the interval between $K = 3$ and $K = 10$. In this interval we also see a monotonic increase in the variance of the CV loss estimation. Based

(2019) due to our inclusion of the whole sample.

[26] In this regard, we also note that in Online Appendix B of Bruhin et al. (2019), the authors estimate $p_{\hat{\theta}_5^*}$ in a specification in which u_5^A has undergone a CES utility transformation. Herein, they find that the parameter estimate of the curvature of the indifference curves is very close to one, indicating that a linear specification, as the one above, is close to optimal in this setting. In turn this indicates that improvements most likely do not come from non-linear utility specifications, but rather by considering non-linear altruism.

on our selection criterion (i.e. the “one-standard-error” rule), it is clear that the optimal number of types in each of the models is $K = 3$.[27]

4.4.2.1 Within-type completeness

Having selected the optimal number of types for the three parametric models, we can now show the estimated within-type completeness of the heterogeneous parametric models. To do this, Table 4.2 shows, for each of the models, indicated in the left most column, an estimation of the within-type completeness for each of the three types nested in the models. As mentioned, the naive benchmark within each type is the same model used as the naive benchmark on the aggregate level, but estimated on the subjects which are assigned to that type. Furthermore, the optimal model is a gradient boosting classifier using the type assignment of each individual as a feature. Finally, the table also shows the estimated size of each type.

Table 4.2: Within-type completeness of heterogeneous parametric models

Model	k	$\hat{N}_k$	$\hat{e}(\ell(\hat{p}_{\hat{\theta}_0^*}))$	$\hat{e}(\ell(\cdot))$	$\hat{e}(\ell(\hat{p}^*))$	$\hat{\kappa}(\cdot)$
$p_{\hat{\theta}_2^k}$	1	77	0.4030	0.2402	0.2191	88.53%
	2	70	0.1862	0.1668	0.1542	60.50%
	3	27	0.5833	0.5191	0.5328	> 100%
$p_{\hat{\theta}_4^k}$	1	76	0.4039	0.2325	0.2147	90.59%
	2	70	0.1862	0.1657	0.1548	65.11%
	3	28	0.5775	0.5174	0.5366	> 100%
$p_{\hat{\theta}_5^k}$	1	78	0.4148	0.2502	0.2381	93.11%
	2	73	0.1870	0.1660	0.1552	65.98%
	3	23	0.5946	0.5040	0.5231	> 100%

Number of subjects: 174
Number of observations: 20,358

Note: $p_{\hat{\theta}_i^k}$ for $k \in \{1,2,3\}$ and $i \in \{2,4,5\}$ are parametric models contained in the mixture models $p_{\hat{\theta}_{i,3}}$ for $i \in \{2,4,5\}$ as defined in Section 4.2.4. k is the type indicator and $\hat{N}_k$ is the size of type k based on the estimated ex post type membership probabilities. $\hat{e}(\ell(\hat{p}_{\hat{\theta}_0^*}))$ is the average negative log likelihood on the test set of the naive benchmark model. $\hat{e}(\ell(\hat{f}^*))$ is the average negative log likelihood on the test set of the optimal model. $\hat{e}(\ell(\cdot))$ is the average negative log likelihood on the test set of respective models. $\hat{\kappa}(\cdot)$ is the estimated completeness on the test set.

We can see that all three models mostly agree on the sizes of the types. In particular,

[27] We note that the CV loss is not in fact minimized at $K = 3$ for any of the models. However, the CV loss estimate at $K = 3$ is within 1/2 of a standard error of the minimum CV loss in all of the models. Thus, given the increasing noise in the estimate and our preference for parsimony, we find $K = 3$ a reasonable selection.

in all models, there are two relative large types and one small. In addition to this, we see that the expected loss of the naive benchmark is very similar across models in the same type. Specifically, this estimated loss for $k = 1$ is between 0.4030 and 0.4148, and we see the same pattern for $k = 2$ and $k = 3$. The same is the case for the expected loss of the optimal mapping. In turn this tells us that the type composition is very similar across models. Specifically, if a given subject belongs to type $k = 1$ in one of these models, then it is very likely the case that she also belongs to type $k = 1$ in any of the others. Based on this, we also see a similar pattern in our estimations of within-type completeness. In particular, all of the models show high completeness in $k = 1$, with $p_{\hat{\theta}^1_5}$ being most complete (93.11%). Hence, within this type, including negative reciprocity increases completeness by about 2 percentage points and including both negative and positive reciprocity increases completeness by an additional approximately 3 percentage points. This shows that within this type (i) differentiated altruism is the by far most important motive, (ii) the partial impact of negative reciprocity is slightly smaller than on the aggregate level, but (iii) the partial impact of positive reciprocity is of comparable size as in the aggregate level. The expected loss of the naive benchmark is quite high, indicating that there is substantial variation within this type. That is, the predictive performance of a parametric model that does not include any other-regarding preferences is quite low. However, the models are able to pick up the vast majority of the predictable patterns. Based on this, we conclude that a simple linear social preference theory can capture the choices of this type very well.

In $k = 2$, we see that choices are relatively easily predictable by using the agents' own payoffs, based on the expected loss of the naive benchmark, and that the models are only able to pick up some of the remaining predictable patterns, resulting in within-type completeness estimates of 60.50% for $p_{\hat{\theta}^2_2}$, 65.11% for $p_{\hat{\theta}^2_4}$ and 65.98% for $p_{\hat{\theta}^2_5}$. Hence, different from the preceding type, positive reciprocity seems to hardly play a role here, whereas only including negative reciprocity leads to an increase in completeness of approximately 5 percentage points. In addition, due to the relative low completeness of all models within this type, we have an indication that subjects of this type potentially use a more complex model of social preference, that may include non-linearity that we do not consider in these models.

Finally, for $k = 3$ we get unexpected results. The expected loss of the naive benchmark for all of the models within this type reveals that choices are very random, in the sense that they are not well predicted using only the agent's own payoff. However, we see that the expected loss of each of the models outperform that of the estimated optimal mapping.[28] The most likely reason for this is a power issue. That is, our ML model is unable to pick up the predictable patterns within this type because (i) choices are very random and (ii) the type consists of a relatively small number of individuals. Based on this, we are unable to evaluate whether the models perform well in terms of prediction within this type. We thus turn to the parameter estimates to see whether these are stable.

Table 4.3: Parameter estimates of heterogeneous parametric models

Model	k	$1/\hat{\sigma}$	$\hat{\gamma_A}$	$\hat{\gamma_D}$	$\hat{\gamma_K}$	$\hat{\gamma_U}$
$p_{\hat{\theta}_2^k}$	1	0.0168^{***}	0.4929^{***}	0.1914^{***}	–	–
		(0.0008)	(0.0206)	(0.0179)		
	2	0.0288^{***}	0.1180^{***}	0.0496^{***}	–	–
		(0.0020)	(0.0145)	(0.0100)		
	3	0.0044^{***}	-0.3194^{*}	-0.8593^{***}	–	–
		(0.0008)	(0.1721)	(0.2327)		
$p_{\hat{\theta}_4^k}$	1	0.0171^{***}	0.5332^{***}	0.2306^{***}	–	-0.1222^{***}
		(0.0009)	(0.0226)	(0.0216)		(0.0213)
	2	0.0289^{***}	0.1282^{***}	0.0599^{***}	–	-0.0312^{***}
		(0.0019)	(0.0161)	(0.0108)		(0.0099)
	3	0.0045^{***}	-0.2072	-0.7434^{***}	–	-0.3422^{**}
		(0.0008)	(0.1656)	(0.2190)		(0.1374)
$p_{\hat{\theta}_5^k}$	1	0.0175^{***}	0.4561^{***}	0.1480^{***}	0.1587^{***}	-0.0451^{**}
		(0.0008)	(0.0235)	(0.0236)	(0.0240)	(0.0206)
	2	0.0289^{***}	0.1290^{***}	0.0622^{***}	-0.0022	-0.0314^{***}
		(0.0018)	(0.0154)	(0.0121)	(0.0114)	(0.0112)
	3	0.0045^{***}	-0.3152^{*}	-0.8511^{***}	0.2005	-0.2415
		(0.0007)	(0.1698)	(0.1938)	(0.1727)	(0.1677)

Number of subjects: 174
Number of observations: 20,358

Note: $p_{\hat{\theta}_2^k}, p_{\hat{\theta}_4^k}, p_{\hat{\theta}_5^k}$ for $k \in \{1, 2, 3\}$ are parametric models contained in mixture models as defined in Section 4.2.4. k is the type indicator. Standard errors clustered on the individual level in parentheses. Significance levels; $^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$.

[28] Note that this result is robust to fitting a distinct gradient boosting classifier within each type of all three models. In addition, due to the small sample size within this type, it may be that a simpler ML model fitted only on this type would be able to outperform the parametric models, although we did not find such a model. However, this would only give us a crude upper bound estimation. The optimal way to deal with this issue, in our opinion, would be to at least double the sample size, such that we have enough observations to properly estimate the irreducible error. This is consistent with findings of Peterson et al. (2021) showing that theory driven models can outperform ML models on smaller data sets due to theory-driven efficiency.

Table 4.3 shows the within-type parameters of each of the heterogeneous parametric models, where the types are matched to those in Table 4.2. Not surprisingly, we see that there is a positive correlation between the predictability within types and the estimated choice sensitive for that type. In particular, in type $k = 2$, where choices are the least random, we also see the highest choice sensitivity in all the models. The estimated choice sensitivity is by far the lowest in $k = 3$ in all the models, which is also the type in which choices, in general, are hard to predict.

From the parameter estimates in Table 4.3, we see that the first type, i.e. $k = 1$, is characterized by substantial altruism in all of the models. The parameter estimate of altruism when ahead, $\hat{\gamma}_A$, varies between 0.4561 and 0.5332 with statistical significance at the 1%-level in all models, indicating a willingness to give up approximate 50 cents to increase the counterpart's payoff by 1 dollar. The parameter estimate of altruism when behind is substantially smaller, but still larger than the estimates on the aggregate level and significant at the 1%-level. Furthermore, we see that, for this type, the model that includes negative reciprocity, $\hat{\gamma}_U$, but not positive reciprocity, $p_{\hat{\theta}^1_4}$, reveals that negative reciprocity has a small impact. However, through the altruism parameter estimates in this model, we see that positive reciprocity plays a much larger role for this type. Only including negative reciprocity leads to a sizeable increase in the altruism parameters, $\hat{\gamma}_A$ and $\hat{\gamma}_D$, as we also saw on the aggregate level, although to a lesser degree. Finally, $p_{\hat{\theta}^1_5}$ reveals the impact of positive reciprocity, $\hat{\gamma}_K$. Specifically, the parameter estimate is significant at the 1%-level and the magnitude is comparable to that of altruism when behind, $\hat{\gamma}_D$. Thus, conditional on a kind preceding action of the counterpart, the agent is willing to sacrifice approximately double as much when behind. On the other hand, negative reciprocity, $\hat{\gamma}_U$, has a much smaller, but statistically significant impact on altruism, which is of comparable magnitude to the estimates on the aggregate level. Hence, from our within-type completeness estimation and parameter estimation of this type, we can conclude that (i) it is very well explained by a linear social preference model and (ii) it is characterized by substantial other-regarding preferences.

In $k = 2$, we see substantially less altruism compared to the previous type. In particular, the parameter estimate of altruism when ahead, $\hat{\gamma}_A$, is between 0.1180 and 0.1290, but significant at the 1%-level in all the models. The parameter estimate of altruism when behind, $\hat{\gamma}_D$, is also significant at the 1%-level for all models and varies between 0.0496 and 0.0622. These estimates are thus closer to those on the aggregate level than in the previous type. Finally, considering $p_{\hat{\theta}_4^2}$ and $p_{\hat{\theta}_5^2}$, we see that negative reciprocity has a small, but statistically significant impact on utility, whereas positive reciprocity seemingly plays no role. Based on our parameter estimates, we can conclude that other-regarding preferences are much less important to this type than the previous. However, the within-type completion estimate also reveals that the other-regarding preferences that are present, are potentially used in a non-linear manner that is not captured by our simple preference models. Hence, for this type, it might be worth exploring other functional forms that allow for concave or convex altruism.

Finally, in the last type, $k = 3$, we see quite counter-intuitive estimates. In particular, the parameter estimate of $\hat{\gamma}_A$ varies between -0.3152 and -0.2072, indicating a willingness to sacrifice a substantial amount to decrease the counterpart's payoff when ahead. However, the estimate is only significant in $p_{\hat{\theta}_2^3}$ and $p_{\hat{\theta}_5^3}$, and only at the 10%-level, indicating a high variance. For the estimates of $\hat{\gamma}_D$, we see substantial behindness aversion with point estimates varying between -0.7434 and -0.8593. Notice also that these are all significant at the 1%-level. Finally, we see that the parameter estimates of positive and negative reciprocity, $\hat{\gamma}_K$ and $\hat{\gamma}_U$, are substantial and have the expected signs. However, due to the large variance, only negative reciprocity is statistically significant and only in $p_{\hat{\theta}_4^3}$. Based on these parameter estimates, we have an indication that a type exists that exhibits severe behindness aversion and malice. Additionally, the type also seems to be driven by both positive and negative reciprocity. However, due to the small sample size and the magnitude of the parameters, we seemingly do not have enough power to fully estimate the behavioral characteristics of this type. This is in line with the within-type completeness estimate showing that the ML model is unable to pick up a substantial part of the predictable patterns on the small sample. We conclude that we would need to include substantially more subjects to get stable and reliable estimates for this type and to evaluate whether a linear social preference model is appropriate for explaining their choices. However, including more subjects may also complicate the estimations. Whereas the two larger types, $k = 1$ and $k = 2$, appear stable, it might be that the smaller type, $k = 3$, consist of two or more minority types that would appear in a larger sample size.

4.4.2.2 Unrestricted completeness

Table 4.7 in the Appendix shows the results of the estimation of our gradient boosting classifier in four distinct variants. The first variant, $p^*_{GB^{ind}}$, that uses the subject identifier has a lower expect loss estimate than the corresponding model on the aggregate level, which ignores the identifier. However, the loss estimate is substantially higher than that of any of the heterogeneous parametric models with three types. This indicates that we do not have a sufficient amount of observations of each subject for the ML model to properly capture the individual characteristics related to predicting the choice.[29] Considering the three variants of the gradient boosting classifier that use a clustering algorithm as a preprocessing step, $p^*_{GB^{km}}, p^*_{GB^{hc}}$ and $p^*_{GB^{ber}}$, we see that the ML model using the K-means algorithm is most successful in terms of the lowest expected loss estimate. Hence, for our estimation of the unrestricted completeness, we will use the expected loss of $p^*_{GB^{km}}$ as our estimate of the optimal mapping, $\hat{p}^*$. However, we would like to stress once again that these estimations should be seen as upper bounds. We do not claim that we have found the absolute minimum loss.

Table 4.4 shows our unrestricted completeness estimates of the three heterogenous parametric models that we consider, $p_{\hat{\theta}^*_{2,3}}$, $p_{\hat{\theta}^*_{4,3}}$ and $p_{\hat{\theta}^*_{5,3}}$. Here the naive benchmark model, $p_{\hat{\theta}_0}$, is identical to the one on the aggregate level, and once again, the completeness of the naive benchmark and the optimal mapping, $\hat{p}^*$, are 0% and 100% by definition, respectively.

As can be seen, the expected loss estimate of all of the heterogeneous models are substantially lower than that of the naive benchmark model, and of that of the respective parametric models on the aggregate level (see Table 4.1). In addition, we see that all of the models are relatively complete, in the sense that they are much closer to the expected loss of the optimal mapping than to that of the naive benchmark. Specifically, the model including only differentiated altruism, $p_{\hat{\theta}^*_{2,3}}$, achieves an estimated completeness of 84.82%, whereas the models that include negative reciprocity, $p_{\hat{\theta}^*_{4,3}}$, and positive and negative reciprocity, $p_{\hat{\theta}^*_{5,3}}$, achieves a completeness estimate of 86.42% and 88.36%, respectively. Hence, we conclude that (i) differentiated altruism is the by far most important other-regarding

[29] A similar result can be derived by fitting the parametric models on each individual separately. Here the expected loss would be minimized by $p_{\hat{\theta}_1}$, indicating a lack of power to fully estimate the parameters of the individuals.

Table 4.4: Unrestricted completeness of heterogeneous parametric models

Model	$\hat{e}(\ell(\cdot))$	$\hat{\kappa}(\cdot)$
$p_{\hat{\theta}_0^*}$	0.3862	0%
$p_{\hat{\theta}_{2,3}^*}$	0.2541	84.82%
$p_{\hat{\theta}_{4,3}^*}$	0.2516	86.42%
$p_{\hat{\theta}_{5,3}^*}$	0.2486	88.36%
$\hat{p}^*$	0.2305	100%
Number of subjects: 174		
Number of observations: 20,358		

Note: $p_{\hat{\theta}_0}$ is a parametric model and $p_{\hat{\theta}_{2,3}^*}$,$p_{\hat{\theta}_{4,3}^*}$,$p_{\hat{\theta}_{5,3}^*}$ are mixture models as defined in Section 4.2.4. $\hat{p}^*$ is a gradient boosting classifier. $\hat{e}(\ell(\cdot))$ is the average negative log likelihood on the test set. $\hat{\kappa}(\cdot)$ is the estimated completeness on the test set.

behavioral motive for predicting the subjects' choice, (ii) a parsimonious representation of the subjects in three types is able to capture most of the individual variation in the data, and (iii) there seems to be little predictive potential in considering models with more types and with non-linear social preferences.

4.5 Discussion

We now shortly discuss the validity and generalizability of our results. The former refers to our assumptions and estimation strategy. The latter refers to the extent that our results on this limited domain can say something about the predictability of the social preference models generally.

The strictest assumption that we impose is that utility noise follows an identical Gumbel distribution. This is a "convenience" assumption making the translation of the social preference models into models that predict the probability of choosing a given allocation straightforward. Whereas this assumption is standardly imposed, it may, naturally, be violated and the extent of the impact on our results will most likely depend on the degree of the violation. In this regard, we note a recent contribution by Alós-Ferrer et al. (2021) showing that preferences of individuals can be recovered by imposing assumptions on decision time rather than the utility noise. Hence, collecting decision time data may be a way to overcome this "convenience" assumption in future work on this domain.

The next point that may influence the validity is our data-splitting strategy. The reasons for using our approach are as follows. Firstly, it provides maximal power in estimating the completeness of the models on both the aggregate and heterogeneous level since each individual is present in both the training and test set with the approximately same number of observations. This makes it more powerful than an approach in which we make our splits completely random. Secondly, it makes it possible, on the heterogeneous level, to construct models that depend on the subject identifier. This allows us to (i) consider very flexible ML models that should be able to capture most of the predictable variation in the data, and (ii) compare social preference models with an increasing number of types. Alternatively, one could split the data in a way that holds out a fixed random subset of the games for all subjects. However, this would result in the prediction problem being not only out-of-sample, but also outside of the observed features in the training stage. As our goal is to estimate the completeness conditional on the features, this is not an optimal approach in this setting. However, such a strategy does not come without value. In fact, such predictions would tell us how well a given model generalizes to settings from which it has not "learned". Finally, a viable alternative approach would be to split the data in a way that holds out a fixed random subset of the individuals. This is indeed an intuitive approach if one wishes to estimate how well a given (potential heterogeneous) model generalizes to the population. However, this approach also introduces several issues that have to be dealt with. For instance, it is not clear how to assign out-of-sample individuals to types as those assignments would be based on the out-of-sample choices. One approach would be to assign out-of-sample individuals to the most prevalent type. However, intuitively, this should lead to even worse performance than aggregate estimation. An alternative approach would be to assume that a random subset of the out-of-sample individual's decisions are known to us. In this way, one could use the information provided by the known choices of the out-of-sample individuals to assign them to types and then use the corresponding models to make predictions on the unknown choices. However, then one might argue that these known choices should be in the training stage of the model in the first place and then we arrive at an approach similar to the one that we are taking.

Finally, with regard to the generalizability of our results, we note that our results may naturally be limited to (i) the domain (i.e., that of dictator games and reciprocity games), (ii) the complexity of the games, and (iii) the number of counterparts that the subject faces. Naturally, to fully grasp the completeness of the considered theories, we should conduct the investigation on more games, in more complex settings (e.g., where reciprocity enters non-binary), and with varying number of counterparts. As our investigation indicates, this should be done with a much larger sample size, such that within-type completeness estimates are possible for all types. An additional benefit of increasing the sample size is that it makes it possible to see whether individual characteristics, such as those collected in this data set (see Table 4.5), can predict the type assignment of the subjects. If this indeed is the case, then that would provide us with information on which characteristics the estimated types consists of. This is an exercise that unfortunately is not possible here.

4.6 Conclusion

This paper extends the literature on theory evaluation to that of social preferences and proposes ways in which parameterzed theories can be evaluated when allowing for heterogeneity. Our results suggest several interesting behavioral patterns on the considered domain. Firstly, on the aggregate level, we find that a linear social preference model including altruism that may depend on the payoff allocations as well as positive and negative reciprocity captures most of the predictable patterns in the data. In particular, we find that such a model is approximately 82% complete, so that potential improvement by considering more complex functional forms is rather limited. When allowing for heterogeneity we see the emergence of three distinct types. The two first types are of similar proportion and significantly larger than the third minority type. We find that the first of the larger types is characterized by strong other-regarding preferences, and our within-type completeness measure indicates that the preferences of this type can be very well predicted by a linear social preference model. The second of the large types is characterized by modest social preferences, but based on our within-type completeness estimations, we find that the behavioral motives most likely interact in a more complex functional form. For the minority type, we find that we are unable to estimate the within-type completeness due to the type's small sample size. However, the parameter estimates indicate that the type is characterized by strong behindness aversion and even inequity loving behavior when ahead. Finally, our unrestricted completeness estimations indicate that a linear social preference theory, as long as it admits differentiated altruism, with only three types is able to capture most of the individual variation in the data. Our results are naturally limited in several ways. Firstly, the domain of binary dictator games and reciprocity games contains only a small subset of situations in which social preferences play a role. In addition, these situations may contain more than the single counterpart that we include in this paper. Thus, to fully grasp the completeness of social preference theories, this analysis should be conducted on more games with varying numbers of counterparts and include other aspects such as beliefs. We note, however, as this analysis shows, that the sample should be sufficiently large to be able to estimate the completeness of all types once allowing for heterogeneity.

4.7 Bibliography

Alós-Ferrer, C., Fehr, E., and Netzer, N. (2021). Time will tell: Recovering preferences when choices are noisy. *Journal of Political Economy*, 129(6):1828–1877.

Bolton, G. E. and Ockenfels, A. (2000). Erc: A theory of equity, reciprocity, and competition. *American economic review*, 90(1):166–193.

Bruhin, A., Fehr, E., and Schunk, D. (2019). The many faces of human sociality: Uncovering the distribution and stability of social preferences. *Journal of the European Economic Association*, 17(4):1025–1069.

Charness, G. and Rabin, M. (2002). Understanding social preferences with simple tests. *The quarterly journal of economics*, 117(3):817–869.

Chi, J. T., Chi, E. C., and Baraniuk, R. G. (2016). k-pod: A method for k-means clustering of missing data. *The American Statistician*, 70(1):91–99.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22.

Fehr, E. and Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The quarterly journal of economics*, 114(3):817–868.

Friedman, J., Hastie, T., and Tibshirani, R. (2009). *The elements of statistical learning*, volume 2. Springer series in statistics New York.

Fudenberg, D., Gao, W., and Liang, A. (2020). How flexible is that functional form? quantifying the restrictiveness of theories. *arXiv preprint arXiv:2007.09213*.

Fudenberg, D., Kleinberg, J., Liang, A., and Mullainathan, S. (2021). Measuring the completeness of economic models. Working paper, MIT Economics.

Fudenberg, D. and Puri, I. (2021). Evaluating and extending theories of choice under risk. Working paper, MIT Economics.

Levine, D. K. (1998). Modeling altruism and spitefulness in experiments. *Review of economic dynamics*, 1(3):593–622.

McLachlan, G. J., Lee, S. X., and Rathnayake, S. I. (2019). Finite mixture models. *Annual review of statistics and its application*, 6:355–378.

Noti, G., Levi, E., Kolumbus, Y., and Daniely, A. (2016). Behavior-based machine-learning: A hybrid approach for predicting human decision making. *arXiv preprint arXiv:1611.10228*.

Peel, D. and MacLahlan, G. (2000). Finite mixture models. *John & Sons*.

Peterson, J. C., Bourgin, D. D., Agrawal, M., Reichman, D., and Griffiths, T. L. (2021). Using large-scale experiments and machine learning to discover theories of human decision-making. *Science*, 372(6547):1209–1214.

Peysakhovich, A. and Naecker, J. (2017). Using methods from machine learning to evaluate behavioral models of choice under risk and ambiguity. *Journal of Economic Behavior & Organization*, 133:373–384.

Plonsky, O., Apel, R., Ert, E., Tennenholtz, M., Bourgin, D., Peterson, J. C., Reichman, D., Griffiths, T. L., Russell, S. J., Carter, E. C., et al. (2019). Predicting human decisions with behavioral theories and machine learning. *arXiv preprint arXiv:1904.06866*.

Plonsky, O., Erev, I., Hazan, T., and Tennenholtz, M. (2017). Psychological forest: Predicting human behavior. In *Thirty-first AAAI conference on artificial intelligence*.

Smyth, P. (2000). Model selection for probabilistic clustering using cross-validated likelihood. *Statistics and computing*, 10(1):63–72.

Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge university press.

4.8 Appendix

Table 4.5: Descriptive statistics of the subjects

Variable	Description	Mean
Age	–	21.71 (3.02)
Income	Monthly (in CHF)	640.17 (491.37)
Female	Binary indicator	0.53 –
Natural Sciences	Binary indicator	0.60 –
Medical Sciences	Binary indicator	0.14 –
Social Sciences	Binary indicator	0.10 –
Law	Binary indicator	0.07 –
Cognitive ability	–	7.01 (2.39)
Consciousness	Big 5 measure	6.98 (3.10)
Openness	Big 5 measure	20.92 (3.85)
Extraversion	Big 5 measure	6.07 (3.93)
Agreeableness	Big 5 measure	8.07 (2.79)
Neuroticism	Big 5 measure	3.88 (4.16)
Number of subjects: 174		

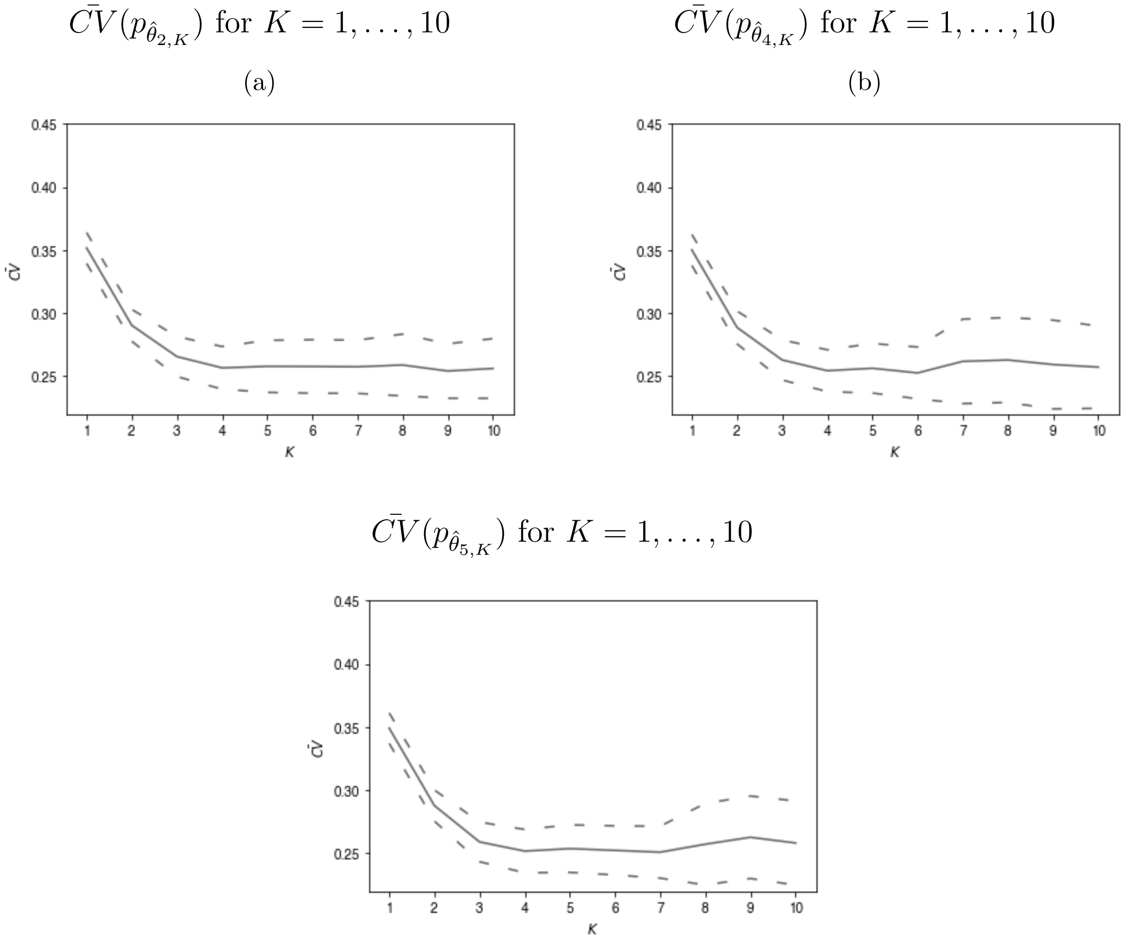
Note: Standard deviation in parentheses.

Table 4.6: Non-parametric and ML benchmarks on the aggregate level

Model	$\bar{CV}(\cdot)$	$\hat{e}(\ell(\cdot))$
$\hat{p}^*_{TL}$	–	0.3651
$\hat{p}^*_{RF}$	0.3384 (0.0126)	0.3348
$\hat{p}^*_{GB}$	0.3374 (0.0127)	0.3345
Number of subjects: 174 Number of observations: 20,358		

Note: $\hat{e}(\ell(\cdot))$ is the average negative log likelihood on the test set of the model. $\bar{CV}(\cdot)$ is the average negative log likelihood of the model with its optimal hyperparameters averaged over the five CV folds. Standard error of the cross validation estimation in parentheses.

Figure 4.2: CV loss of heterogeneous parametric models with varying number of types



$\bar{CV}(p_{\hat{\theta}_{2,K}})$ for $K = 1, \ldots, 10$

(a)

$\bar{CV}(p_{\hat{\theta}_{4,K}})$ for $K = 1, \ldots, 10$

(b)

$\bar{CV}(p_{\hat{\theta}_{5,K}})$ for $K = 1, \ldots, 10$

(c)

Note: In all plots the solid line is the average negative log likelihood averaged over the 5 estimates in the cross validation procedure for the finite mixture model with K types. The dashed line is +/- one standard error of this estimate.

Table 4.7: ML benchmarks on the heterogeneous level

Model	$\bar{CV}(\cdot)$	$\hat{e}(\ell(\cdot))$
$\hat{p}^*_{GB^{ind}}$	0.3165 (0.0159)	0.3081
$\hat{p}^*_{GB^{km}}$	0.2344 (0.0143)	0.2305
$\hat{p}^*_{GB^{hc}}$	0.2698 (0.0105)	0.2434
$\hat{p}^*_{GB^{ber}}$	0.2563 (0.0180)	0.2363

Number of subjects: 174
Number of observations: 20,358

Note: $\hat{e}(\ell(\cdot))$ is the average negative log likelihood on the test set of the model. $\bar{CV}(\cdot)$ is the average negative log likelihood of the model with its optimal hyperparameters averaged over the five CV folds. Standard error of the cross validation estimation in parentheses.

Appendix to dissertation

Declaration of contribution to the chapters of the dissertation

My co-author Christopher Kops and I estimated my respective contributions to Chapter 2 and 3 together.

This or that? - Sequential rationalization of indecisive choice behavior (Chapter 2)

Contribution:	
Research idea:	50%
Analysis:	50%
Development of manuscript:	50%

Managing anticipations and reference-dependent choice (Chapter 3)

Contribution:	
Research idea:	50%
Analysis:	50%
Development of manuscript:	50%

Evaluating the completeness of social preference theories (Chapter 4)

Contribution:	
Data acquisition:	0%
Research idea:	100%
Analysis:	100%
Development of manuscript:	100%

Curriculum vitae

Personal Information

Name	Jesper Armouti-Hansen
Address	Merkenicher Str. 28 50735 Cologne Germany
Date of Birth	13.04.1986 in Herlev, Denmark
Citizenship	Danish
Email	jeshan49@gmail.com

Employment

12/2015 –	Teaching and Research Assistant at the Seminar of Personnel Economics and HRM, Prof. Dr. Dirk Sliwka, University of Cologne

Education

11/2015 –	Doctoral Student at the University of Cologne,
10/2012 – 03/2015	M.Sc. of International Economics and Public Policy, Johannes Gutenberg University of Mainz
08/2008 – 01/2012	B.A. of Financial Management and Services, Copenhagen Business Academy (Denmark)
08/2003 – 07/2006	Herlev Gymnasium (Denmark)

Published Papers

Cassar, Lea, and Jesper Armouti-Hansen (2020): “Optimal contracting with endogenous project mission.” *Journal of the European Economic Association*, 18(5), pp. 2647–2676

Armouti-Hansen, Jesper, and Christopher Kops (2018): “This or that? - Sequential rationalization of indecisive choice behavior” *Theory and Decision*, 84(4), pp. 507–524.

Refereeing

Scandinavian Journal of Economics, Journal of Economic Behavior and Organization

Eidesstattliche Erklärung

Hiermit versichere ich an Eides Statt, dass ich die vorgelegte Dissertation selbstständig und ohne die Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen direkt oder indirekt übernommenen Aussagen, Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet. Bei der Auswahl und Auswertung folgenden Materials haben mir die nachstehend aufgeführten Personen in der jeweils beschriebenen Weise unentgeltlich geholfen: /

Weitere Personen neben den in der Einleitung der Dissertation aufgeführten Koautorinnen und Koautoren waren an der inhaltlich-materiellen Erstellung der vorliegenden Dissertation nicht beteiligt. Insbesondere habe ich hierfür nicht die entgeltliche Hilfe von Vermittlungs- bzw. Beratungsdiensten in Anspruch genommen. Niemand hat von mir unmittelbar oder mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen. Die Dissertation wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt. Ich versichere, dass ich nach bestem Wissen die reine Wahrheit gesagt und nichts verschwiegen habe.

Köln, 20.10.2021